



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :

C12N 15/12, 15/85, C07K 14/47, 16/18,  
C12Q 1/68, G01N 33/53, A61K 38/17,  
48/00

A2

(11) International Publication Number:

WO 97/41226

(43) International Publication Date:

6 November 1997 (06.11.97)

(21) International Application Number:

PCT/US97/07172

(22) International Filing Date:

28 April 1997 (28.04.97)

(30) Priority Data:

60/016,465

29 April 1996 (29.04.96)

US

(71) Applicant (for all designated States except US): THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by the SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Bethesda, MD 20892 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): JOHNSON, Alfred, C. [US/US]; 3829 Wendy Lane, Silver Spring, MD 20906 (US).

(74) Agents: STORELLA, John, R. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: POLYNUCLEOTIDES ENCODING THE TRANSCRIPTIONAL REPRESSOR GCF2 OF THE EPIDERMAL GROWTH FACTOR RECEPTOR

(57) Abstract

The invention provides GCF2 protein or GCF2 protein analogs that inhibit transcription from the EGFR and other promoters. The invention also provides polynucleotides encoding the polypeptides, antibodies against the polypeptides and methods of using them.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

5 POLYNUCLEOTIDES ENCODING THE TRANSCRIPTIONAL REPRESSOR GCF2 OF THE EPIDERMAL GROWTH FACTOR RECEPTOR

BACKGROUND OF THE INVENTION

10 This invention relates to the field of molecular biology, including binding proteins, their recombinant production, and therapeutic use.

The epidermal growth factor receptor ("EGFR") plays an important role in cell growth and development (Carpenter, G. (1987) "Receptors for epidermal growth factor and other polypeptide mitogens." *Biochem.* 56:881-914; Hernandez-Sotomayor, S.M., and G. Carpenter (1992) "Epidermal growth factor receptor: elements of intracellular communication." *J. Membr. Biol.* 128:81-89; Merlino, G.T. (1990) "Epidermal growth factor receptor regulation and function." *Semin. Cancer Biol.* 1:277-284). Over-expression of the EGFR can lead to epidermal growth factor-dependent transformation (DiFiore, P.P. et al. (1987) "Overexpression of the human EGF receptor confers an EGF-dependent transformed phenotype to NIH 3T3 cells." *Cell* 51:1063-1070; Velu, T. J. et al. (1987) "Epidermal-growth-factor-dependent transformation by a human EGF receptor proto-oncogene." *Science* 238:1408-1410). Over-production of EGFR has been detected in several types of cancers due to gene amplification (King, C.R. et al (1985) "Human tumor cell lines with EGF receptor gene amplification in the absence of aberrant sized mRNAs." *Nucleic Acids Res.* 13:8477-8486). Over-expression of EGFR transcripts in a variety of other tumors such as ovarian, cervical and kidney tumors results from transcriptional or post-transcriptional mechanisms (Xu, Y.H., et al. (1984) "Characterization of epidermal growth factor receptor gene expression in malignant and normal human cell lines." *Proc. Natl. Acad. Sci USA* 81:7308-7312).

30 A variety of agents have been shown to increase EGFR gene expression (Hou, X., et al. (1994) "Induction of epidermal growth factor receptor gene transcription by transforming growth factor beta 1: association with loss of protein binding to a negative regulatory element." *Cell Growth Differ.* 5:801-809; Hudson, L.G. and G.N.

Gill (1991) "Regulation of gene expression by epidermal growth factor." *Genet. Eng.* 13:137-151; Hudson, L.G. et al. (1990) "Identification and characterization of a regulated promoter element in the epidermal growth factor receptor gene." *Proc. Natl. Acad. Sci. USA* 87:7536-7540). Repression of EGFR gene transcription by different agents has also been reported (Hudson, L.G. (1990) "Ligand-activated thyroid hormone and retinoic acid receptors inhibit growth factor receptor promoter expression." *Cell* 62:1165-1175; Zheng, Z.S., (1992) "Transcriptional control of epidermal growth factor receptor by retinoic acid." *Cell Growth Differ.* 3:225-232). Transcriptional control plays a major role in regulation of EGFR gene expression.

The promoter of the EGFR gene lacks a "TATA box" and "CAAT box" but contains multiple "GC boxes" and multiple transcription initiation sites. A number of regions in the promoter have been identified that bind nuclear factors (Chen, L.L. et al. (1993) "A sequence-specific single-stranded DNA-binding protein that is responsive to epidermal growth factor recognizes and S1 nuclease-sensitive region in the epidermal growth factor receptor promoter." *Cell Growth Differ.* 4:975-983; Johnson, A.C. et al. (1988) "Epidermal growth factor receptor gene promoter. Deletion analysis and identification of nuclear protein binding sites." *J. Biol. Chem.* 263:5693-5699; Johnson, A.C. et al. (1988) "Modulation of epidermal growth factor receptor proto-oncogene transcription by a promoter site sensitive to S1 nuclease." *Mol. Cell Biol.* 8:4174-4184). Furthermore, Sp1, wild type p53 and ETF have been shown to activate EGFR gene transcription (Deb, S.P., et al. (1994) "Wild-type human p53 activates the human epidermal growth factor receptor promoter." *Oncogene* 9:1341-1349; Kageyama, R., et al. (1988) "Epidermal growth factor (EGF) receptor gene transcription. Requirement for Sp1 and an EGF receptor-specific factor." *J. Biol. Chem.* 263:6329-6336; Kageyama, R. (1989) "Nuclear factor ETF specifically stimulates transcription from promoters without a TATA box." *J. Biol. Chem.* 264:15508-15514). Two repressor proteins, ETR (EGFR transcriptional repressor and GC (GC-binding factor) also bind to sites within the EGFR promoter (Hou, X. et al. (1994) "Identification of an epidermal growth factor receptor transcriptional repressor." *J. Biol. Chem.* 269:4307-4312; Kageyama, R. (1989) "Nuclear factor ETF specifically stimulates transcription from promoters without a TATA box." *J. Biol. Chem.* 264:15508-15514).

The cDNA for GCF1 was isolated by screening an A431 expression library with GC-rich sequences from the EGFR promoter (Kageyama, R., and I. Pastan

(1989) "Molecular cloning and characterization of a human DNA binding factor that represses transcription." *Cell* 59:815-825). GCF1 is a 91 kDa protein that binds to three upstream sites of the EGFR promoter. Two are between -270 and -225 bp and the other site is between -150 and -90 relative to the translational start site. Cotransfection experiments have shown that GCF1 can repress transcription of the EGFR promoter and several other growth related gene promoters such as transforming growth factor  $\alpha$  (TGF- $\alpha$ ) and insulin like growth factor II (Kitadai, Y. et al. (1993) "GC factor represses transcription of several growth factor/receptor genes and causes growth inhibition of human gastric carcinoma cell lines." *Cell Growth Differ.* 4:291-296). The cDNA for GCF1 hybridizes to three mRNA species of 4.5, 3.0 and 1.2 kb (Johnson, A.C. et al. "Expression and chromosomal localization of the gene for the human transcriptional repressor GCF." *J. Biol. Chem.* 267:1689-1694). The GCF1 cDNA is 2.8 kb in size and is likely to be derived from the 2.0 kb mRNA.

## SUMMARY OF THE INVENTION

A cDNA encoding a new transcription repressor protein, GCF2, has been discovered. This protein represses transcription from the epidermal growth factor receptor (EGFR) promoter, the SV40 promoter and Rous sarcoma virus (RSV) promoter. The ability of GCF2 to repress EGFR expression is important because EGFR expression is increased in certain cancers, e.g., breast cancer.

GCF2 mRNA is expressed in most human tissues as a 4.2 kb mRNA with high level expression in peripheral blood leukocytes. GCF2 mRNA is predominantly expressed in heart and skeletal muscle as a 2.9 kb mRNA. Also, most normal tissue have an additional hybridizing species of 2.4 kilobases.

Cancer cell lines do not express the 2.4 kb species, or do so only very weakly. High levels of GCF2 are found in breast cancer and B and T cell lymphomas. Also, high levels of a GCF2 are expressed in Raji cells and HUT cells. Furthermore, GCF2 binding to the EGFR promoter is reduced in breast cancer cells.

The gene for GCF2 is localized to chromosome 20q13.3.

Accordingly, this invention is directed to purified GCF2 protein whose amino acid sequence is substantially identical to the amino acid sequence of SEQ ID NO:2. The invention also is directed to GCF2 protein analogs whose amino acid sequence is not naturally occurring and which comprises a contiguous sequence of at

least 10 amino acids from the amino acid sequence of native GCF2 (SEQ ID NO:2). In one embodiment, the analog, when presented as an immunogen, elicits the production of an antibody which specifically binds to native GCF2 protein. In another embodiment, the GCF2 protein analog binds to the EGFR promoter and/or inhibits the expression of a nucleotide sequence operably linked to the EGFR gene promoter.

In another aspect, this invention is directed to recombinant polynucleotides comprising a nucleotide sequence of at least 25 contiguous nucleotides from nucleotides 128 to 1384 or 1694 to 2310 of SEQ ID NO:1. In one embodiment, this invention provides a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence that codes for the expression of a polypeptide whose amino acid sequence comprises a contiguous sequence of at least 10 amino acids selected from amino acids 1 to 752 of SEQ ID NO:2. In another aspect, the invention provides recombinant host cells transfected with an expression vector comprising a recombinant polynucleotide having expression control sequences operably linked with a sequences that codes for the expression of a polypeptide having a sequence of at least 10 amino acids selected from amino acids 1 to 752 of SEQ ID NO:2.

Methods of producing a GCF2 protein or GCF2 protein analog can involve transfecting a host cell with an expression vector having expression control sequences operably linked to nucleotide sequences encoding the protein or peptide analog, and culturing the recombinant cell.

This invention also is directed to isolated polynucleotide probes comprising at least 15 nucleotides, that specifically hybridize with a unique nucleotide sequence of native GCF2 cDNA, SEQ ID NO:1 or with its complement.

In another aspect, this invention is directed to compositions comprising an antibody that specifically binds native GCF2 protein.

In another aspect, this invention is directed to methods for detecting GCF2 binding activity in a sample. The methods involve contacting the sample with a GCF2 binding substrate, and detecting the presence of a GCF2 protein/substrate bound complex. The presence of the complex indicates the presence of GCF2 binding activity in the sample. The method can be a diagnostic method using tumor cells or malignant cells from a subject. In one embodiment, the method is quantitative for determining the amount of GCF2 binding activity in a sample. The method involves determining the

amount of bound complex and comparing the amount with a standard amount of complex based on known amounts of native GCF2 protein.

5 In another aspect, this invention provides kits useful for detecting a bound complex between a GCF2 protein and the GCF2 binding substrate. The kits include a GCF2 binding substrate and an anti-GCF2 antibody. In another embodiment, the kit comprises a GCF2 binding substrate and a GCF2 protein or GCF2 protein analog having DNA binding activity useful as a standard.

10 In another aspect, this invention provides methods for isolating DNA sequences that bind to a GCF2 protein comprising contacting a GCF2 protein or a GCF2 protein analog having GCF2 binding activity with a DNA library.

15 In another aspect, this invention provides *in vitro* methods for inhibiting transcription of a nucleotide sequence operably linked to a promoter regulated by GCF2 comprising providing a GCF2 protein or active GCF2 protein analog to the cell. In one embodiment, the protein is provided by expressing a GCF2 protein or an active GCF2 protein analog in a recombinant host cell from a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence coding for the expression of a polypeptide whose amino acid sequence comprises a contiguous sequence of at least 10 amino acids selected from amino acids 1 to 752 of SEQ ID NO:2 and wherein the polypeptide inhibits the expression of a nucleotide sequence operably linked to the EGFR gene promoter, the RSV promoter or the SV40 promoter.

20 In another aspect, the invention provides methods for restoring GCF2 binding activity in a cancer cell that exhibits reduced GCF2 binding activity in a subject, or inhibiting the growth of cancer cells that over-express the EGF receptor in a subject. The method comprises providing the cell with a GCF2 protein or active GCF2 protein analog. In one embodiment, the protein is provided by transfecting the cancer cell with a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence coding for the expression of a polypeptide whose amino acid sequence comprises a contiguous sequence of at least 10 amino acids selected from amino acids 1 to 752 of SEQ ID NO:2 and wherein the polypeptide inhibits the expression of a nucleotide sequence operably linked to the EGFR gene promoter, the RSV promoter or the SV40 promoter.

30 In another aspect, this invention provides a method of detecting GCF2 mRNA or cDNA in a sample. The method comprises the steps of: (a) contacting the

sample with a probe or primer of this invention; and (b) detecting specific hybridization of the probe or primer to GCF2. Specific hybridization provides a detection of GCF2 mRNA or cDNA in the sample.

5 In another aspect, this invention provides a method for aiding in the diagnosis of cancer. The method comprises the steps of (a) determining a diagnostic value by detecting one or more GCF2 mRNA species in a patient sample; and (b) comparing the diagnostic value with a normal range of the species in a control cell sample. A diagnostic value that is above the normal range is diagnostic of cancer. In one embodiment, the mRNA species are about 4.2 kb and about 2.4 kb. Diagnostic values of the 4.2 kb species above the normal range or values of the 2.4 kb species below the normal range provides a positive sign in the diagnosis of cancer. In one embodiment the cancer is breast cancer, a B-cell lymphoma or a T-cell lymphoma.

10 In another aspect, this invention provides a method of detecting a chromosomal translocation of a GCF2 gene comprising the steps of (a) hybridizing a labeled probe of the invention to a chromosome spread from a cell sample to determine the pattern of hybridization and (b) determining whether the pattern of hybridization differs from a normal pattern. A translocation at this site can result in alteration of GCF2 activity, such as activated transcription or changed function.

15 In another aspect, this invention provides a method of detecting polymorphic forms of GCF2 comprising comparing the identity of a nucleotide or amino acid at a selected position from the sequence of a test GCF2 gene or polypeptide with identity of the nucleotide or amino acid at the corresponding position of native GCF2. A difference in identity indicates that the test polynucleotide is a polymorphic form of GCF2.

25

### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. The nucleotide sequence of native GCF2 cDNA (SEQ ID NO:1) with open reading frame ATG and terminator codon TAA underlined.

30 Figure 2. Deduced amino acid sequence of native GCF2 protein (SEQ ID NO:2). The open reading frame of the GCF2 cDNA was translated into protein sequence using MacVector to generate the 752 amino acids. Underlined sequences represent potential phosphorylation sites, dotted sequence represent a potential N-glycosylation site and asterisks represent a putative nuclear localization signal.



Figure 3. Schematic representation of GCF2 cDNA clones and RACE product. Depicted are the two largest cDNA clones and the 5'RACE product. The schematic is drawn to scale.

Figures 4A-4B. Northern blot analysis with GCF1 cDNA fragments.

5 Fragments containing GCF1 cDNA sequences (A) 1 to 282 and (B) 314 to 961 were labeled and used to probe nitrocellulose filters containing poly (A)<sup>+</sup> RNA from A431 cells (lane 1) and KB cell (lane 2). Filters were processed as described in Materials and Methods and exposed to film at -80°C for 12 hours. RNA sizes were estimated based on migration of ribosomal RNAs.

10 Figure 5. Homology of GCF2 and GCF1 cDNAs. GCF1 and GCF2 were aligned using default parameters for the BestFit sequence analysis software package of the Genetics Computer Group (GCG). Numbers to the left and right of the sequences represent the respective nucleotides of the cDNAs.

15 Figure 6. Northern blot analysis with GCF2 cDNA fragment. Total RNA from A431 cells (lane 1), KB cells (lane 2) and HUT102 cells (lane 3) were transferred to nitrocellulose and probed with a 1.1 kilobase pair GCF2 fragment. The size of the hybridizing RNA was determined by comparison to an RNA ladder (Life Technologies).

20 Figure 7. *In vitro* translation production of GCF2 in rabbit reticulocyte lysates. GCF2 and luciferase were synthesized in the presence of <sup>35</sup>S methionine as described in the Examples. Translated products were analyzed on a 6% SDS polyacrylamide gel. After processing, the dried gel was exposed to film at -80°C for 4 hours.

25 Figure 8. Purification of bacterially expressed GCF2-His. GCF2 was expressed as a His-Tag fusion protein upon IPTG induction of JM109 cells containing pGCF2-His. Sonicates were prepared and GCF2-His purified by nickel affinity chromatography. Samples before and after affinity chromatography were subjected to analysis on a 6% SDS polyacrylamide gel. The gel was fixed and stained with coomassie blue. Lanes: 1) Molecular weight markers; 2) Total soluble fraction;  
30 3) Pooled eluted fractions from nickel affinity column.

Figures 9A-9C. Gel mobility shift assay with GCF2-His and EGFR promoter fragments. EGFR promoter fragments were end-labeled and incubated with GCF2-His as described in the Examples. Samples were analyzed on a 4%

and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

"Allelic variant" refers to any of two or more polymorphic forms of a gene occupying the same genetic locus. Allelic variations arise naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. "Allelic variant" also refers to polymorphisms in non-coding sequences at a genetic locus and cDNAs derived from mRNA transcripts of genetic allelic variants, as well as the proteins encoded by them.

"Hybridizing specifically to" or "specific hybridization" or "selectively hybridize to," refers to the binding, duplexing, or hybridizing of a polynucleotide preferentially to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA.

"Stringent conditions" refers to conditions under which a probe will hybridize preferentially to its target subsequence, and to a lesser extent to, or not at all to, other sequences. "Stringent hybridization" and "stringent hybridization wash conditions" in the context of polynucleotide hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of polynucleotides is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5° C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T<sub>m</sub> for a particular probe.

An example of stringent hybridization conditions for hybridization of complementary polynucleotides which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42° C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15 M NaCl at 72° C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65° C for 15 minutes (see, Sambrook et al. for a

description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45° C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100  
5 nucleotides, is 4-6x SSC at 40° C for 15 minutes. In general, a signal-to-noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

A first sequence is an "antisense sequence" with respect to a second sequence if a polynucleotide whose sequence is the first sequence specifically hybridizes  
10 with a polynucleotide whose sequence is the second sequence.

"Primer" refers to a polynucleotide that is capable of specifically hybridizing to a designated polynucleotide template and providing a point of initiation for synthesis of a complementary polynucleotide. Such synthesis occurs when the polynucleotide primer is placed under conditions in which synthesis is induced, i.e., in  
15 the presence of nucleotides, a complementary polynucleotide template, and an agent for polymerization such as DNA polymerase. A primer is typically single-stranded, but may be double-stranded. Primers are typically deoxyribonucleic acids, but a wide variety of synthetic and naturally occurring primers are useful for many applications. A primer is complementary to the template to which it is designed to hybridize to serve as a site for  
20 the initiation of synthesis, but need not reflect the exact sequence of the template. In such a case, specific hybridization of the primer to the template depends on the stringency of the hybridization conditions. Primers can be labeled with, e.g., chromogenic, radioactive, or fluorescent moieties and used as detectable moieties.

"Probe" refers to a polynucleotide that is capable of specifically  
25 hybridizing to a designated sequence of another polynucleotide. A probe specifically hybridizes to a target complementary polynucleotide, but need not reflect the exact complementary sequence of the template. In such a case, specific hybridization of the probe to the target depends on the stringency of the hybridization conditions. Probes can be labeled with, e.g., chromogenic, radioactive, or fluorescent moieties and used as  
30 detectable moieties.

"Detecting" refers to determining the presence, absence, or amount of an analyte in a sample, and can include quantifying the amount of the analyte in a sample or per cell in a sample.

"Detectable moiety" or a "label" refers to a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include  $^{32}\text{P}$ ,  $^{35}\text{S}$ , fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin-streptavidin, dioxigenin, haptens and proteins for which antisera or monoclonal antibodies are available, or polynucleotides with a sequence complementary to a target. The detectable moiety often generates a measurable signal, such as a radioactive, chromogenic, or fluorescent signal, that can be used to quantitate the amount of bound detectable moiety in a sample. The detectable moiety can be incorporated in or attached to a primer or probe either covalently, or through ionic, van der Waals or hydrogen bonds, e.g., incorporation of radioactive nucleotides, or biotinylated nucleotides that are recognized by streptavidin. The detectable moiety may be directly or indirectly detectable. Indirect detection can involve the binding of a second directly or indirectly detectable moiety to the detectable moiety. For example, the detectable moiety can be the ligand of a binding partner, such as biotin, which is a binding partner for streptavidin, or a nucleotide sequence, which is the binding partner for a complementary sequence, to which it can specifically hybridize. The binding partner may itself be directly detectable, for example, an antibody may be itself labeled with a fluorescent molecule. The binding partner also may be indirectly detectable, for example, a polynucleotide having a complementary nucleotide sequence can be a part of a branched DNA molecule that is in turn detectable through hybridization with other labeled polynucleotides. Quantitation of the signal is achieved by, e.g., scintillation counting, densitometry, or flow cytometry.

"Linker" refers to a molecule that joins two other molecules, either covalently, or through ionic, van der Waals or hydrogen bonds, e.g., a polynucleotide that hybridizes to one complementary sequence at the 5' end and to another complementary sequence at the 3' end, thus joining two non-complementary sequences.

"Amplification" refers to any means by which a polynucleotide sequence is copied and thus expanded into a larger number of polynucleotides, e.g., by reverse transcription, polymerase chain reaction, and ligase chain reaction.

"Polypeptide" refers to a polymer composed of amino acid residues, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof linked via peptide bonds, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof. Synthetic polypeptides can be

synthesized, for example, using an automated polypeptide synthesizer. The term "protein" typically refers to large polypeptides. The term "peptide" typically refers to short polypeptides.

5 Conventional notation is used herein to portray polypeptide sequences: the left-hand end of a polypeptide sequence is the amino-terminus; the right-hand end of a polypeptide sequence is the carboxyl-terminus.

Terms used to describe sequence relationships between two or more nucleotide sequences or amino acid sequences include "reference sequence," "selected from," "comparison window," "identical," "percentage of sequence identity,"  
10 "substantially identical," "complementary," and "substantially complementary."

A "reference sequence" is a defined sequence used as a basis for a sequence comparison and may be a subset of a larger sequence, e.g., a complete cDNA, protein, or gene sequence.

15 Because two polynucleotides or polypeptides each may comprise (1) a sequence (i.e., only a portion of the complete polynucleotide or polypeptide sequence) that is similar between the two polynucleotides, or (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions  
20 of sequence similarity.

A "comparison window" refers to a conceptual segment of typically at least 12 consecutive nucleotide or 4 consecutive amino acid residues that is compared to a reference sequence. The comparison window frequently is at least 15 or at least 25 nucleotides in length or at least 5 or at least 8 amino acids in length. The comparison  
25 window may comprise additions or deletions (i.e., gaps) of about 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerized implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package  
30 Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI) or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by any of the various methods is selected.

A subject nucleotide sequence or amino acid sequence is "identical" to a reference sequence if the two sequences are the same when aligned for maximum correspondence over the length of the nucleotide or amino acid sequence.

5 The "percentage of sequence identity" between two sequences is calculated by comparing two optimally aligned sequences over a comparison window, determining the number of positions at which the identical nucleotide or amino acid occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. 10 Unless otherwise specified, the comparison window used to compare two sequences is the length of the shorter sequence.

When percentage of sequence identity is used in reference to polypeptides it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to known algorithm. See, e.g., Meyers & Miller (1988) 25 *Computer Applic. Biol. Sci.* 4:11-17; Smith & Waterman (1981) *Adv. Appl. Math.* 2:482; Needleman & Wunsch (1970) *J. Mol. Biol.* 48:443; Pearson & Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85:2444; Higgins & Sharp (1988) *Gene* 73:237-244; Higgins & Sharp, *CABIOS* 5:151-153 (1989); Corpet et al. (1988) *Nucleic Acids Research* 16:10881-90; Huang et al. (1992) *Computer Applications in the Biosciences* 8:155-65; 30 and Pearson et al. (1994) *Methods in Molecular Biology* 24:307-31. Alignment is also often performed by inspection and manual alignment.

A subject nucleotide sequence or amino acid sequence is "substantially identical" to a reference sequence if the subject amino acid sequence or nucleotide

sequence has at least 80% sequence identity over a comparison window. Thus, sequences that have at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 98% sequence identity or at least 99% sequence identity with the reference sequence are also "substantially identical." Two sequences  
5 that are identical to each other are, of course, also "substantially identical".

"Complementary" refers to the topological compatibility or matching together of interacting surfaces of two polynucleotides. Thus, the two molecules can be described as complementary, and furthermore, the contact surface characteristics are complementary to each other. A first polynucleotide is complementary to a second  
10 polynucleotide if the nucleotide sequence of the first polynucleotide is identical to the nucleotide sequence of the polynucleotide binding partner of the second polynucleotide. Thus, the polynucleotide whose sequence 5'-TATAC-3' is complementary to a polynucleotide whose sequence is 5'-GTATA-3'.

A nucleotide sequence is "substantially complementary" to a reference  
15 nucleotide sequence if the sequence complementary to the subject nucleotide sequence is substantially identical to the reference nucleotide sequence.

"Conservative substitution" refers to the substitution in a polypeptide of an amino acid with a functionally similar amino acid. The following six groups each contain amino acids that are conservative substitutions for one another:

- 20       1) Alanine (A), Serine (S), Threonine (T);  
         2) Aspartic acid (D), Glutamic acid (E);  
         3) Asparagine (N), Glutamine (Q);  
         4) Arginine (R), Lysine (K);  
         5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and  
25       6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

"Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well  
30 as the myriad immunoglobulin variable region genes. Antibodies exist, e.g., as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. This includes, e.g., Fab' and F(ab)'; fragments. The term "antibody," as used herein, also includes antibody fragments either produced by the

modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies.

5 An antibody "specifically binds to" or "is specifically immunoreactive with" a protein when the antibody functions in a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind preferentially to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. 10 A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay 15 formats and conditions that can be used to determine specific immunoreactivity.

"Immunoassay" refers to an assay that utilizes an antibody to specifically bind an analyte. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the analyte.

20 "Substantially pure" means an object species is the predominant species present (i.e., on a molar basis, more abundant than any other individual organic biomolecular species in the composition), and a substantially purified fraction is a composition wherein the object species comprises at least about 50% (on a molar basis) of all organic biomolecular species present. Generally, a substantially pure composition means that about 80% to 90% or more of the organic biomolecular species present in the 25 composition is the purified species of interest. The object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) if the composition consists essentially of a single organic biomolecular species. "Organic biomolecule" refers to an organic molecule of biological origin, e.g., proteins, polynucleotides, carbohydrates or lipids. Solvent species, small 30 molecules (< 500 Daltons), stabilizers (e.g., BSA), and elemental ion species are not considered organic biomolecular species for purposes of this definition.

"Naturally-occurring" as applied to an object refers to the fact that the object can be found in nature. For example, a polypeptide or polynucleotide sequence



that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

5 "Pharmaceutical composition" refers to a composition suitable for pharmaceutical use in a mammal. A pharmaceutical composition comprises a pharmacologically effective amount of an active agent and a pharmaceutically acceptable carrier. "Pharmacologically effective amount" refers to that amount of an agent effective to produce the intended pharmacological result. "Pharmaceutically acceptable carrier" refers to any of the standard pharmaceutical carriers, buffers, and excipients, such as a  
10 phosphate buffered saline solution, 5% aqueous solution of dextrose, and emulsions, such as an oil/water or water/oil emulsion, and various types of wetting agents and/or adjuvants. Suitable pharmaceutical carriers and formulations are described in *Remington's Pharmaceutical Sciences*, 19th Ed. (Mack Publishing Co., Easton, 1995). Preferred pharmaceutical carriers depend upon the intended mode of administration of the  
15 active agent. Typical modes of administration include enteral (e.g., oral) or parenteral (e.g., subcutaneous, intramuscular, or intravenous intraperitoneal injection; or topical, transdermal, or transmucosal administration).

A "subject" of diagnosis or treatment is an animal, such as a mammal, including a human. Non-human animals subject to treatment include, for example, fish,  
20 birds, and mammals such as cows, sheep, pigs, horses, dogs and cats.

A "prophylactic" treatment is a treatment administered to a subject who does not exhibit signs of a disease or exhibits only early signs for the purpose of decreasing the risk of developing pathology.

A "therapeutic" treatment is a treatment administered to a subject who  
25 exhibits signs of pathology for the purpose of diminishing or eliminating those signs.

"Prognostic value" refers to an amount of an analyte in a subject sample that is consistent with a particular prognosis for a designated disease. The amount (including a zero amount) of the analyte detected in a sample is compared to the prognostic value for the sample such that the relative comparison of the values indicates  
30 the likely outcome of the progression of the disease.

"Diagnostic value" refers to a value that is determined for an analyte in a subject sample, which is then compared to a normal range of the analyte in a sample (e.g., from a healthy individual) such that the relative comparison of the values provides

a reference value for diagnosing a designated disease. Depending upon the method of detection, the diagnostic value may be a determination of the amount of the analyte, but it is not necessarily an amount. The diagnostic value may also be a relative value, such as a plus or a minus score, and also includes a value indicating the presence or absence of the analyte in a sample.

## II. GCF2 PROTEINS

This invention provides purified GCF2 proteins having an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:2. In one embodiment a "GCF2 protein" is native GCF2, whose amino acid sequence is identical to the amino acid sequence of SEQ ID NO:2. Native GCF2 protein has no significant amino acid homology with any other known protein. In another embodiment, a "GCF2 protein" is a human allelic variant or an animal cognate of native GCF2 that can be encoded by a polynucleotide that hybridizes under stringent conditions to the nucleotide sequence encoding native GCF2 of SEQ ID NO:1 and that is isolatable from human or animal cDNA or genomic libraries. Thus, GCF2 proteins have a naturally occurring (i.e., existing in nature) amino acid sequence.

This invention also provides GCF2 protein analogs. As used herein, the term "GCF2 protein analog" refers to a non-naturally occurring polypeptide comprising a contiguous sequence of at least 10 amino acids, at least 15 amino acids, at least 20 amino acids or at least 25 amino acids from the sequence of native GCF2 (SEQ ID NO:2). In one embodiment, GCF2 protein analogs, when presented as an immunogen, elicit the production of an antibody which specifically binds to native GCF2 protein. GCF2 protein analogs optionally are in isolated form.

This invention also provides active GCF2 protein analogs that bind the EGFR promoter, as determined by gel mobility shifts when the protein is incubated with a DNA fragment containing the promoter (e.g., SEQ ID NO:3 or 4) and that inhibit the expression of nucleotide sequences operably linked to the EGFR gene promoter. An active analog inhibits expression of a sequence operably linked to the EGFR promoter if the amount of transcription of an mRNA from the sequence is decreased in a statistically significant amount (usually at least 5-fold or at least 10-fold) in recombinant host cells that express the analog, compared with the amount of transcription from cells that do not express the analog. It is expected that an active GCF2 protein analog will comprise an

amino acid sequence substantially identical to the lysine-rich sequence of amino acids 511 to 524 of SEQ ID NO:2.

Active GCF2 analogs preferably have a contiguous sequence of at least 550 acids substantially identical to an amino acid sequence, either contiguous or non-contiguous, from the 752 amino-acid protein of native GCF2 or, more preferably, have a contiguous sequence of 675 amino acids having at least 95% sequence identity with an amino acid sequence, either contiguous or non-contiguous, from native GCF2. It is expected that active GCF2 protein analogs will include a sequence having substantial identity to at least amino acids 511 to 524 of native GCF2.

Active GCF2 protein analogs include GCF2 protein analogs whose amino acid sequence differs from that of native GCF2 by the inclusion of amino acid substitutions, additions or deletions (e.g., active fragments). Active fragments can be identified empirically by cutting back the protein from either the amino-terminus or the carboxy-terminus to generate fragments, and testing the resulting fragments for activity.

Active analogs bearing substitutions can be prepared by introducing conservative amino acid substitutions into the native protein. The number of substitutions is at the discretion of the practitioner, but the amino acid sequence of the resulting protein must conform to the definition of active GCF2 protein analogs, above.

Active GCF2 protein analogs having additions include those having amino acid extensions to the amino- or carboxy-terminal end of other active fragments, as well as additions made internally to the protein. In one embodiment, terminal amino acid sequences are added encoding a polyhistidine tag to simplify purification.

GCF2 protein analogs that are oligopeptides can be prepared by chemical synthesis using well known methods. However, both oligopeptides and larger GCF2 proteins and protein analogs preferably are prepared recombinantly.

### III. POLYNUCLEOTIDES

A cDNA molecule encoding a GCF2 protein and portions of the untranslated 5' and 3' regions has been isolated. The nucleotide sequence and deduced amino acid sequence of the polynucleotide are presented in Figures 1 and 2. (SEQ ID NO:1 and SEQ ID NO:2, respectively.) This nucleotide sequence contains an open reading frame of 2256 bases encoding native GCF2 protein from nucleotide 125 to

nucleotide 2380. The GCF2 cDNA hybridizes to a mRNA of 4.2 kb from HUT102, A431 and KB cells. The expression is highest in HUT102 cells. The polypeptide having the deduced amino acid sequence has a calculated molecular weight of 83 kDa protein in SDS-PAGE. The bacterially expressed protein binds selectively to EGFR promoter fragments having the sequence 5' CGGGCAGCCC CCGGCGC 3' (SEQ ID NO:3).  
5 Cotransfection assays show that GCF2 acts to repress transcription from the EGFR, SV40 and RSV (Rous sarcoma virus) promoters.

The GCF2 nucleotide sequence of SEQ ID NO:1 contains a 309 base pair sequence from nucleotide 1385 to 1693 that has 98% homology, 305/309 base pairs, to  
10 the GCF1 cDNA. (Fig. 2, SEQ ID NO: 4.) The initiation codon of GCF1 begins at nucleotide 224 of SEQ ID NO:4, which is homologous to position 1608 of SEQ ID NO:1. Because GCF1 is translated in a different reading frame than GCF2, the GCF1 nucleotide sequence does not code for the expression of the same amino acid sequence. However, because both nucleotide sequences are rich in adenine residues around  
15 nucleotides 1653 to 1693 of SEQ ID NO:1 (nucleotides 271 to 309 of SEQ ID NO:4), they both encode a region rich in lysine residues.

Accordingly, this invention provides recombinant polynucleotides comprising nucleotide sequences from the sequence encoding GCF2 protein, and nucleotide sequences that code for the expression of a GCF2 protein or GCF2 protein  
20 analog. In one embodiment, the recombinant polynucleotide comprises at least 25, at least 30, at least 50, at least 100, at least 500 or at least 1000 nucleotides in a contiguous sequence from nucleotide 128 (just after the initiation codon) to nucleotide 1384 (just before the area of homology with GCF1) or nucleotide 1694 (just after the area of homology with GCF1) to nucleotide 2310 (just before the termination codon) of SEQ ID  
25 NO:1.

In another embodiment, the recombinant polynucleotide comprises a nucleotide sequence that codes for the expression of a polypeptide whose amino acid sequence comprises a contiguous sequence of at least 10 amino acids, at least 15 amino acids, at least 20 amino acids, at least 25 amino acids, at least 100 amino acids or at  
30 least 500 amino acids from amino acids 1 to 752 of SEQ ID NO:2. In one alternative, at least 10 amino acids are selected from amino acids 1 to 461 or 562 to 752 of SEQ ID NO:2 i.e., outside the area encoded by the nucleotide sequence having a high degree of homology with GCF1. In one embodiment, the nucleotide sequence is substantially

identical to the nucleotide sequence of SEQ ID NO:1. In another embodiment, the nucleotide sequence that encodes the contiguous amino acid sequence (e.g., at least 10 amino acids) selected from amino acids 1 to 752 of SEQ ID NO:2 is a nucleotide sequence from SEQ ID NO:1.

5 In another embodiment, the recombinant polynucleotide is an expression vehicle. In this embodiment, the recombinant polynucleotide can comprise expression control sequences operably linked to a nucleotide sequence encoding at least 10 amino acids selected from amino acids 1 to 752 of SEQ ID NO:2.

10 In one embodiment, the nucleotide sequence codes for the expression of a protein which, when presented as an immunogen, elicits the production of an antibody which specifically binds to native GCF2 protein. The nucleotide sequence also can code for the expression of a protein that inhibits the expression of a nucleotide sequence operably linked to the EGFR gene promoter, the SV40 promoter or the RSV promoter. Preferably, such a protein contains an amino acid sequence substantially identical to  
15 amino acids 511 to 524 of SEQ ID NO:2.

In another embodiment, the nucleotide sequence coding for expression of a GCF2 protein or GCF2 protein analog has a contiguous sequence of 1650 nucleotides substantially identical to a nucleotide sequence, either contiguous or non-contiguous, from the 2256 nucleotide sequence encoding native GCF2 of SEQ ID NO:1 or, more  
20 preferably, has a contiguous sequence of 2025 nucleotides having at least 95% sequence identity with a nucleotide sequence, either contiguous or non-contiguous, from that sequence.

The polynucleotides of the present invention are cloned, or amplified by *in vitro* methods, such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), the self-sustained sequence  
25 replication system (3SR) and the Q $\beta$  replicase amplification system (QB). For example, a polynucleotide encoding the protein can be isolated by polymerase chain reaction of cDNA from HUT102, A431 or KB cells using primers based on the DNA sequence of GCF2 of SEQ ID NO:1. A wide variety of cloning and *in vitro* amplification  
30 methodologies are well-known to persons of skill. PCR methods are described in, for example, U.S. Pat. No. 4,683,195; Mullis et al. (1987) *Cold Spring Harbor Symp. Quant. Biol.* 51:263; and Erlich, ed., *PCR Technology*, (Stockton Press, NY, 1989). Polynucleotides also can be isolated by screening genomic or cDNA libraries with probes

selected from the sequences of SEQ ID NO:1 under stringent hybridization conditions, e.g., salt and temperature conditions substantially equivalent to 5x SSC and 65°C for both hybridization and wash.

5 Mutant versions of the proteins can be made by site-specific mutagenesis of other polynucleotides encoding the proteins, or by random mutagenesis caused by increasing the error rate of PCR of the original polynucleotide with 0.1 mM MnCl<sub>2</sub> and unbalanced nucleotide concentrations.

10 This invention also provides expression vectors, e.g., recombinant polynucleotides further comprising expression control sequences operatively linked to the nucleotide sequence coding for expression of the polypeptide. Expression vectors can be adapted for function in prokaryotes or eukaryotes by inclusion of appropriate promoters, replication sequences, markers, etc. The construction of expression vectors and the expression of genes in transfected cells involves the use of molecular cloning techniques also well known in the art. Sambrook et al., *Molecular Cloning -- A Laboratory*  
15 *Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (1989) and *Current Protocols in Molecular Biology*, F.M. Ausubel et al., eds., (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.)

20 Methods of transfecting genes into mammalian cells and obtaining their expression for *in vitro* use or for gene therapy, are well known to the art. See, e.g., *Methods in Enzymology*, vol. 185, Academic Press, Inc., San Diego, CA (D.V. Goeddel, ed.) (1990) or M. Krieger, *Gene Transfer and Expression -- A Laboratory Manual*, Stockton Press, New York, NY, (1990).

25 Expression vectors useful in this invention depend on their intended use. Such expression vectors must, of course, contain expression and replication signals compatible with the host cell. Expression vectors useful for expressing the protein of this invention include viral vectors such as retroviruses, adenoviruses and adeno-associated viruses, plasmid vectors, cosmids, liposomes and the like. Viral and plasmid vectors are preferred for transfecting mammalian cells. The expression vector pcDNA1 (Invitrogen, San Diego, CA), in which the expression control sequence comprises the  
30 CMV promoter, provides good rates of transfection and expression. Adeno-associated viral vectors are useful in the therapeutic methods of this invention. Appropriate expression control sequences for mammalian cells include, for example, the metallothionein promoter and CMV (cytomegalovirus).

The construct can also contain a tag to simplify isolation of the protein. For example, a polyhistidine tag of, e.g., six histidine residues, can be incorporated at the amino terminal end of the fluorescent protein substrate. The polyhistidine tag allows convenient isolation of the protein in a single step by nickel-chelate chromatography.

5           The invention also provides recombinant host cells transfected with the expression vector for expression of the nucleotide sequences coding for expression of a polypeptide of this invention. Host cells can be selected for high levels of expression in order to purify the protein. Mammalian cells are preferred for this purpose, but prokaryotic cells, such as *E. coli*, also are useful. The cell can be, e.g., a cultured cell  
10 or a cell *in vivo*.

This invention is also directed to polynucleotide probes and primers, preferably isolated, of at least 15 nucleotides, at least 20 nucleotides or at least 25 nucleotides, that specifically hybridize with a nucleotide sequence of nucleotide sequence of SEQ ID NO:1 or its complement, in particular, a unique sequence. As used herein  
15 "unique nucleotide sequence of SEQ ID NO:1" refers to nucleotide sequences between nucleotides 1 to 1384 or 1694 to 3523 of SEQ ID NO:1. In one embodiment, the probe has a sequence identical or complementary to a sequence of SEQ ID NO:1. These isolated polynucleotides are useful as primers for amplification of GCF2 sequences by, e.g., PCR. They also are useful as probes in hybridization assays, such as Southern and  
20 Northern blots, for identifying polynucleotides having a nucleotide sequence of a protein of this invention. In one embodiment, the isolated polynucleotides further comprise a label.

#### IV. ANTIBODIES AND HYBRIDOMAS

25           In another embodiment, this invention provides a composition comprising an antibody that specifically binds GCF2 proteins. Preferably, the antibody does not specifically bind GCF1. Antibodies preferably have affinity of at least  $10^6 \text{ M}^{-1}$ ,  $10^7 \text{ M}^{-1}$ ,  $10^8 \text{ M}^{-1}$ , or  $10^9 \text{ M}^{-1}$ .

##### 30           A. Production of Antibodies

A number of immunogens are used to produce antibodies that specifically bind GCF2 polypeptides. Recombinant or synthetic polypeptides of 10 amino acids in length, or greater, selected from sub-sequences of SEQ ID NO:2 are the preferred

polypeptide immunogen for the production of monoclonal or polyclonal antibodies. In one class of preferred embodiments, an immunogenic peptide conjugate is also included as an immunogen. Naturally occurring polypeptides are also used either in pure or impure form.

5 Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide, or a synthetic version thereof, is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies can be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

10 Methods of producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified polypeptide, a polypeptide coupled to an appropriate carrier (e.g., GST, keyhole limpet hemanocyanin, etc.), or a polypeptide incorporated into an immunization vector such as a recombinant vaccinia virus (see, U.S. Patent No. 4,722,848) is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen  
15 preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, e.g., Coligan (1991) *Current Protocols in Immunology*  
20 Wiley/Greene, NY; and Harlow and Lane (1989) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY.

Antibodies, including binding fragments and single chain recombinant versions thereof, against predetermined fragments of GCF2 proteins are raised by immunizing animals, e.g., with conjugates of the fragments with carrier proteins as  
25 described above. Typically, the immunogen of interest is a peptide of at least about 3 amino acids, more typically the peptide is 5 amino acids in length, preferably, the fragment is 10 amino acids in length and more preferably the fragment is 15 amino acids in length or greater. The peptides can be coupled to a carrier protein (e.g., as a fusion protein), or are recombinantly expressed in an immunization vector. Antigenic  
30 determinants on peptides to which antibodies bind are typically 3 to 10 amino acids in length.



Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies are screened for binding to normal or modified polypeptides, or screened for agonistic or antagonistic activity, e.g., activity mediated through GCF2.

5 In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies are found in, e.g., Stites et al. (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane, *Supra*; Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY; and  
10 Kohler and Milstein (1975) *Nature* 256: 495-497. Summarized briefly, this method proceeds by injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are fused with myeloma cells. The result is a hybrid cell or "hybridoma" that is capable of reproducing *in vitro*. The population of hybridomas is then screened to isolate individual clones, each of which secrete a single  
15 antibody species to the immunogen. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods known in the art. Colonies  
20 arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells is enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate (preferably mammalian) host. The polypeptides and antibodies of the present invention are used with or without modification, and include  
25 chimeric antibodies such as humanized murine antibodies.

Other suitable techniques involve selection of libraries of recombinant antibodies in phage or similar vectors. See, Huse et al. (1989) *Science* 246: 1275-1281; and Ward, et al. (1989) *Nature* 341: 544-546.

30 Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties,

chemiluminescent moieties, magnetic particles, and the like. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen et al. (1989) *Proc. Nat'l Acad. Sci. USA* 86: 10029-10033.

The antibodies of this invention are also used for affinity chromatography in isolating GCF2 proteins. Columns are prepared, e.g., with the antibodies linked to a solid support, e.g., particles, such as agarose, Sephadex, or the like, where a cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified GCF2 polypeptides are released.

The antibodies can be used to screen expression libraries for particular expression products such as mammalian GCF2. Usually the antibodies in such a procedure are labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

Antibodies raised against GCF2 can also be used to raise anti-idiotypic antibodies. These are useful for detecting or diagnosing various pathological conditions related to the presence of the respective antigens.

An alternative approach is the generation of humanized immunoglobulins by linking the CDR regions of non-human antibodies to human constant regions by recombinant DNA techniques. See Queen et al., *Proc. Natl. Acad. Sci. USA* 86:10029-10033 (1989) and WO 90/07861. The humanized immunoglobulins have variable region framework residues substantially from a human immunoglobulin (termed an acceptor immunoglobulin) and complementarily determining regions substantially from a mouse immunoglobulin, (referred to as the donor immunoglobulin). The constant region(s), if present, are also substantially from a human immunoglobulin. The human variable domains are usually chosen from human antibodies whose framework sequences exhibit a high degree of sequence identity with the murine variable region domains from which the CDRs were derived. The heavy and light chain variable region framework residues can be derived from the same or different human antibody sequences. The human antibody sequences can be the sequences of naturally occurring human antibodies or can be consensus sequences of several human antibodies. See Carter et al., WO 92/22653. Certain amino acids from the human variable region framework residues are selected for substitution based on their possible influence on CDR conformation and/or binding to

antigen. Investigation of such possible influences is by modeling, examination of the characteristics of the amino acids at particular locations, or empirical observation of the effects of substitution or mutagenesis of particular amino acids.

For example, when an amino acid differs between a murine variable  
5 region framework residue and a selected human variable region framework residue, the human framework amino acid should usually be substituted by the equivalent framework amino acid from the mouse antibody when it is reasonably expected that the amino acid:

- (1) noncovalently binds antigen directly,
- (2) is adjacent to a CDR region,
- 10 (3) otherwise interacts with a CDR region (e.g., is within about 3 Å of a CDR region), or
- (4) participates in the  $V_L$ - $V_H$  interface.

Other candidates for substitution are acceptor human framework amino  
acids that are unusual for a human immunoglobulin at that position. These amino acids  
15 can be substituted with amino acids from the equivalent position of the antibody or from the equivalent positions of more typical human immunoglobulins.

A further approach for isolating DNA sequences which encode a human  
monoclonal antibody or a binding fragment thereof is by screening a DNA library from  
human B cells according to the general protocol outlined by Huse et al., *Science*  
20 246:1275-1281 (1989) and then cloning and amplifying the sequences which encode the  
antibody (or binding fragment) of the desired specificity. The protocol described by  
Huse is rendered more efficient in combination with phage display technology. See,  
e.g., Dower et al., WO 91/17271 and McCafferty et al., WO 92/01047. Phage display  
technology can also be used to mutagenize CDR regions of antibodies previously shown  
25 to have affinity for GCF2 protein receptors or their ligands. Antibodies having improved  
binding affinity are selected.

In another embodiment of the invention, fragments of antibodies against  
GCF2 protein or protein analogs are provided. Typically, these fragments exhibit  
specific binding to the GCF2 protein receptor similar to that of a complete  
30 immunoglobulin. Antibody fragments include separate heavy chains, light chains Fab,  
Fab' F(ab')<sub>2</sub>, Fabc, and Fv. Fragments are produced by recombinant DNA techniques,  
or by enzymic or chemical separation of intact immunoglobulins.

## V. METHODS FOR DETECTING GCF2 POLYNUCLEOTIDES

The probes and primers of this invention are useful, among other things, in detecting GCF2 polynucleotides in a sample. A method for detecting the presence, absence or amount of a GCF2 polynucleotide in a sample involves two steps: (1) specifically hybridizing a polynucleotide probe or primer to a GCF2 polynucleotide, and (2) detecting the specific hybridization.

For the first step of the method, the polynucleotide used for specific hybridization is chosen to hybridize to any suitable region of GCF2. The polynucleotide can be a DNA or RNA molecule, as well as a synthetic, non-naturally occurring analog of the same. The polynucleotides in this step are polynucleotide primers and polynucleotide probes disclosed herein.

For the second step of the reaction, any suitable method for detecting specific hybridization of a polynucleotide to GCF2 may be used. Such methods include, e.g., amplification by extension of a hybridized primer using reverse transcriptase (RT); extension of a hybridized primer using RT-PCR or other methods of amplification; and *in situ* detection of a hybridized primer. In *in situ* hybridization, a sample of tissue or cells is fixed onto a glass slide and permeabilized sufficiently for use with *in situ* hybridization techniques. Detectable moieties used in these methods include, e.g., labeled polynucleotide probes; direct incorporation of label in amplification or RT reactions, and labeled polynucleotide primers.

Often, cell extracts or tissue samples used in methods for determining the amount of a polynucleotide in a sample will contain variable amounts of cells or extraneous extracellular matrix materials. Thus, a method for determining the cell number in a sample is important for determining the relative amount per cell of a test polynucleotide such as GCF2. A control for cell number and amplification efficiency is useful for determining diagnostic values for a sample of a potential cancer, and a control is particularly useful for comparing the amount of test polynucleotide such as GCF2 in sample to a prognostic value for cancer. A preferred embodiment of the control RNA is endogenously expressed 28S rRNA. (See, e.g., Khan et al., (1992) *Neurosci. Lett.* 147:114-117, which used 28S rRNA as a control, by diluting reverse transcribed 28S rRNA and adding it to the amplification reaction.)

## VI. DIAGNOSTIC METHODS

### A. GCF2 Binding To EGFR Promoter

It has been discovered that in nuclear extracts of the breast cancer cell line, MDA-MB-231, the binding activity of GCF2 to the EGFR gene promoter is reduced. Over-expression of the EGF receptor is known to lead to malignant transformation. Because GCF2 inhibits the transcription of the EGFR gene, reduced GCF2 binding activity in these cells is implicated as a link in the chain of events leading to transformation. Therefore, detection of GCF2 binding activity in cancer cells is useful as a diagnostic tool in detecting the malignant state and uncovering its etiology. Detection of GCF2 binding activity also is useful as a research tool in the study of the regulation of transcription.

Accordingly, this invention provides methods for detecting GCF2 binding activity in a sample. The methods involve contacting the sample with a GCF2 binding substrate, and detecting binding between GCF2 and the substrate, in particular, by detecting the presence of bound GCF2 protein/substrate complex in the sample. The amount of binding activity can be determined by determining the amount of complex and comparing it with a standard amount of complex based on known amounts of native GCF2 protein.

The sample preferably is a nuclear extract from the cell to be tested. However, it can be whole cell extract as well. In diagnostic methods for cancer, the sample preferably derives from a malignant cell known to exhibit an increase in EGF receptor expression. This includes, for example, breast cells, ovary cells, cervix cells and kidney cells.

As used herein, a "GCF2 binding substrate" is a polynucleotide to which native GCF2 protein binds. Native GCF2 protein binds to DNA sequences in the EGFR promoter and other promoters as well. For example, the GCF2 binding substrate can be a polynucleotide comprising the nucleotide sequence of SEQ ID NO:3, from the EGFR promoter. Other nucleotide sequences to which native GCF2 binds are also contemplated. Such sequences also can be determined empirically by, for example, probing DNA libraries with native GCF2 to identify sequences to which the protein binds. The substrate can be labelled to enhance detection. To allow optimal binding between GCF2 in the sample and the GCF2 binding substrate, the two are incubated preferably for at least 15 minutes at about room temperature (i.e., about 23 °C).

The presence of GCF2 binding activity in the sample can be determined by detecting the GCF2 protein/substrate bound complex. One means of doing so is by gel mobility shift assay. The complex between GCF2 protein and the binding substrate has greater mass than GCF2 protein, alone. Thus, the presence of binding can be detected by detecting these larger mass complexes. Immunological methods are useful. In one method, the proteins in the sample are separated by SDS PAGE, and GCF2 is detected by probing the gel with anti-GCF2 antibodies. Alternatively, the test DNA molecule can be radioactively labeled, and the complex detected on the gel by autoradiography.

Another method useful in quantitation involves a sandwich assay in which the binding substrate is immobilized on a surface. The sample is contacted with the surface under conditions for binding, unbound molecules are washed away, and the surface is contacted with labelled anti-GCF2 antibodies. The amount of bound label can be measured, and provides a quantitative measurement of the amount of binding.

#### B. GCF2 Levels

It also has been found that levels of GCF2 are increased in certain cancers, such as breast cancer and B-cell and T-cell lymphomas. Also, the 2.4 kb species of GCF2 mRNA that is found in most normal cells is not expressed, or expressed only weakly in cancer cells. These facts are useful in the diagnosis of cancers. According to one method of the invention, one detects the amount of various species of GCF2 mRNA in cancer cells and compares that amount to a normal range. Increased levels of the 4.2 kb species is a positive of cancer. Decreased amounts of the 2.4 kb species also is a positive sign of cancer. Hybridization can be detected by any means known in the art, including RT-PCR or *in situ* hybridization.

#### VII. KITS

This invention also provides kits for performing GCF2 binding activity detection assays. In one embodiment, the kit contains a GCF2 binding substrate and an anti-GCF2 antibody for detecting the bound complex. In another embodiment, the kit contains a GCF2 binding substrate and a GCF2 protein for use as an activity standard. In another embodiment, the kit contains a GCF2 binding substrate, an anti-GCF2 antibody for detecting the bound complex, and a GCF2 protein for use as an activity

standard. In these kits, the substrate, the antibody and/or the antibody can be labeled. The kit also can contain instructions for carrying out the assay.

#### VIII. METHODS OF INHIBITING THE EXPRESSION OF GENES OPERABLY LINKED WITH EGFR PROMOTER

5 The expression vectors of this invention also are useful *in vitro* for studying the control of transcription of genes and for studying the effect of inhibiting the expression of the EGF receptor. Accordingly, this invention provides methods for inhibiting transcription of a gene operably linked to an EGFR gene promoter, an RSV  
10 promoter or a SV40 promoter. The method involves expressing a GCF2 protein or an active GCF2 protein analog in a recombinant host cell from a recombinant polynucleotide comprising a nucleotide sequence that codes for the expression of a GCF2 protein or GCF2 protein analog. The cells can be cells that express the EGF receptor, or cells co-transfected with a gene operably linked to a promoter whose transcription is regulated  
15 by GCF2. For example, the expression cassette can be an EGFR promoter operably linked to a nucleotide sequence encoding the EGF receptor. The promoter can be endogenous, i.e., a native EGFR gene promoter.

#### IX. THERAPEUTIC METHODS

20 Reduced GCF2 binding activity in cancer cells can result in increased expression of the EGF receptor which, in turn, can lead to malignancy. Therefore, restoration of GCF2 binding activity in a malignant cell exhibiting reduced GCF2 binding activity and/or over-expression of the EGF receptor is useful in the treatment of cancer. Accordingly, this invention provides therapeutic methods for restoring GCF2 binding  
25 activity in a cancer cell that exhibits reduced GCF2 binding activity or inhibiting the growth of cancer cells that over-express the EGF receptor in an individual. The methods involve transfecting the cancer cell with a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence coding for the expression of a GCF2 protein or active GCF2 protein analog that inhibits the expression  
30 of a nucleotide sequence operably linked to the EGFR gene promoter. Expression of the protein in the cell inhibits the expression of the EGF receptor. The recombinant polynucleotide can be delivered by any of the known methods for *in vivo* delivery, including those mentioned above.

## X. GENOMICS

The identification of cognate or polymorphic forms of the GCF2 gene and the tracking of those polymorphisms in individuals and families is important in genetic screening. Accordingly, this invention provides methods useful in detecting polymorphic forms of the GCF2 gene. The methods involve comparing the identity of a nucleotide or amino acid at a selected position from the sequence of a test GCF2 gene with the nucleotide or amino acid at the corresponding position from the sequence of native GCF2. The comparison can be carried out by any methods known in the art, including direct sequence comparison by nucleotide sequencing, sequence comparison or determination by hybridization or identification of RFLPs.

In one embodiment, the method involves nucleotide or amino acid sequencing of the entire test polynucleotide or polypeptide, or a subsequence from it, and comparing that sequence with the sequence of native GCF2. In another embodiment, the method involves identifying restriction fragments produced upon restriction enzyme digestion of the test polynucleotide and comparing those fragments with fragments produced by restriction enzyme digestion of native GCF2 gene. Restriction fragments from the native gene can be identified by analysis of the sequence to identify restriction sites. Another embodiment involves the use of oligonucleotide arrays. (See, e.g., Fodor et al., United States patent 5,445,934.) The method involves providing an oligonucleotide array comprising a set of oligonucleotide probes that define sequences selected from the native GCF2 sequence, generating hybridization data by performing a hybridization reaction between the target polynucleotide molecules and the probes in the set and detecting hybridization between the target molecules and each of the probes in the set and processing the hybridization data to determine nucleotide positions at which the identity of the target molecule differs from that of native GCF2. The comparison can be done manually, but is more conveniently done by a programmable, digital computer.

While not wishing to be limited by theory, it is believed that the lack of the 2.4 kb message in cancer cells results from a mutant form of the GCF2 gene. Accordingly, detection of mutant forms of the gene is useful in identifying cells as cancerous or potentially cancerous.



The following examples are offered by way of illustration, not by way of limitation.

### EXAMPLES

#### 5 I. MATERIALS AND METHODS

##### A. Cell Culture

Cells were maintained in medium supplemented with 10% fetal bovine serum (Life Technologies). Medium was removed and cells washed with phosphate-buffered saline without  $\text{Ca}^{++}$  and  $\text{Mg}^{++}$  prior to RNA isolation.

##### 10 B. RNA Isolation And Blotting

Total RNA was isolated by the guanidinium-thiocyanate-phenol-chloroform extraction method of Chomczynski and Sacchi (Chomczynski, P. and N. Sacchi. (1987) "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." *Anal. Biochem.* 162:156-159). Poly(A)<sup>+</sup> RNA was selected from the total RNA population by oligo(dT)-cellulose chromatography (Aviv, H. and P. Leder. (1972) "Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose." *Proc. Natl. Acad. Sci. USA* 69:1402-1412)).

20 Labeled cDNA probes were prepared by random primer extension of PCR generated fragments as described by Feinberg and Vogelstein. Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* 137:266-267. Tissue blots and the cancer cell line blot were purchased from Clontech and probed according to manufacturer's instructions.

##### 25 C. Isolation And Sequence Analysis Of GCF2 cDNA Clones

The 282 bp GCF1 cDNA fragment, nucleotides 1-282 (SEQ ID NO:4), was labeled with ( $\alpha$ -<sup>32</sup>P)dCTP and used as a hybridization probe to screen an ovarian carcinoma (OVCAR-3) cell cDNA library constructed in Uni-Zap XR (Stratagene). Positive clones were purified and phagemids were excised by use of R408 helper phage (Stratagene). The clones were sequenced with Applied Biosystems model 373A automated DNA sequencer. Sequence comparisons were performed with BLAST and PROSITE using the default parameter to search the National Center for Biotechnology Information nonredundant protein and DNA databases (Altschul, S.F. et al. (1990)

"Basic local alignment search tool." *J. Mol. Biol.* 215:403-410; Bairoch, A. (1993) "The PROSITE dictionary of sites and patterns in proteins, its current status." *Nucleic Acids Res.* 21:3097-3103).

5           D.     5' Rapid Amplification Of cDNA Ends (RACE)

5' RACE-Ready cDNA (liver) was purchased from Clontech, Palo Alto, CA. GCF2-specific primers were selected using Oligo 4.0 (National Biosciences). Nested primers were used to enhance specificity. The 5' RACE product detected after primary and secondary amplification was purified by agarose gel electrophoresis, subcloned into pCRII (Invitrogen) and sequenced. The RACE products contained  
10           homology to the GCF2 cDNA clones and extended to the 5' end. The full-length GCF2 cDNA was constructed by ligation of restriction fragments.

          E.     *In vitro* Transcription and Translation

15           The open reading frame of GCF2 was amplified by the polymerase chain reaction (PCR) and subcloned into pCITE2A (Invitrogen) to generate pCITE-GCF2. Protein was synthesized *in vitro* into the presence of (<sup>35</sup>S)-methionine with the coupled transcription translation system (TNT) from Promega Corporation. Translated products were analyzed on SDS-polyacrylamide gels (Laemmli, U.K. (1970) "Cleavage of  
20           structural proteins during the assembly of the head of bacteriophage T4." *Nature* 227:680-685).

          F.     Bacterial Expression And Purification

          The GCF2 open reading frame was cloned into pQE60 (Qiagen) at the  
25           BamHI site after addition of BamHI linkers to the open reading frame by PCR. The new plasmid, pGCF2-His was sequenced to check for mutations and used to transform JM109. JM109 cells containing pGCF2-His were induced with 1 mM IPTG at  $A_{600} = 0.7$  for 4.5 hr. Cells were harvested and resuspended in sonication buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl). Cells were subjected to two cycles  
30           of freezing and thawing followed by treatment with lysozyme (1 mg/ml) for 30 min on ice. The sample was then sonicated (1 min bursts/1 min cooling/200-300 watts) on ice and treated with 10  $\mu$ g/ml RNase A for 15 min. After centrifugation at 10,000 x g for 20 min, the supernatant was mixed with Ni-NTA resin for 60 min at 4°C. The mixture

was loaded into a column and washed with sonication buffer followed by sonication buffer plus 0.8 mM imidazole and sonication buffer plus 40 mM imidazole. The GCF2-His-Tag protein was eluted in sonication buffer plus 0.5 M imidazole and examined by SDS-PAGE. Fractions containing GCF2-His were dialyzed versus a buffer  
5 containing 20 mM HEPES pH 7.9, 20 mM KCl, 1 mM MgCl<sub>2</sub>, 2 mM DTT and 17% glycerol. Dialyzed samples were stored in aliquots at -80°C.

#### G. Gel Mobility Shift Assays

Mobility shift assays were previously described (Johnson, A.C. et al.  
10 (1988) "Epidermal growth factor receptor gene promoter. Deletion analysis and identification of nuclear protein binding sites." *J. Biol. Chem.* 263:5693-5699). Briefly, end-labeled EGF receptor promoter fragments were incubated with GCF2-His at room temperature (23°C) for 15 min in the presence of 10 mM Tris pH 7.5, 1 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl, 50 µg/ml poly(dI-dC) (poly dI-dC) and  
15 4% glycerol. Samples (20 µl) were loaded onto a 5% polyacrylamide gel and subjected to electrophoresis at 150 volts for 2 hr using 0.5 X TBE (1 X TBE = 89 mM Tris, 8 mM boric acid and 2 mM EDTA, pH 8.3) as running buffer. After electrophoresis, gels were transferred to Whatman 3 MM paper and exposed to Kodak XAR film with intensifying screens at -70°C.

#### H. DNase I Footprinting

DNase I footprinting was performed according to Dynan et al. (Dynan, W.S., and R. Tjian (1983) "The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter." *Cell* 35:79-87). The EGF receptor  
25 promoter fragment (-771 to -16) was labeled at the HindIII site and a 553 base pair (-569 to -16) fragment isolated after restriction digestion with TaqI. GCF2-His was prepared as described above.

#### I. Transfections And CAT Assays

30 African green monkey kidney cells (CV-1) were seeded at 5 x 10<sup>5</sup> cells per 100 mm dishes incubated overnight at 37°C in a 5% CO<sub>2</sub> incubator. For each transfection, 2 µg to 10 µg of pCMV-GCF2 and 2 µg of pERCAT6 DNA mixed in 1.4 ml Opti-MEM (Life Technologies) and a precipitate formed using lipofectamine (Life

Technologies) according to manufacturers recommendations. The cells were washed with serum free DMEM and complexes applied to the cells for 5 hrs. DMEM containing 10% fetal bovine serum was added and cells incubated overnight. Media was changed the following day and cells grown for an additional 24 hr. Cells were harvested and extract prepared as described previously (Gorman, C.M. et al. (1982) "Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells." *Mol. Cell Biol.* 2:1044-1051). Chloramphenicol acetyltransferase (CAT) activity was assayed in extracts using the CAT assay kit from Promega Corporation. Transfection efficiency was monitored by measuring beta-galactosidase activity from a RSV- $\beta$ -galactosidase reporter plasmid construct that was also co-transfected.

J. Expression of GST-GCF2 fusion protein in bacteria

BamH1 linkers were ligated onto a GCF2 cDNA(OQ) fragment encoding amino acids 51-705. After linker ligation and BamH1 digestion, the purified fragment was ligated into pGEX1lt (Pharmacia) that was BamH1 cut and dephosphorylated. An aliquot of the ligation was used to transform JM109 that were plated on LB-ampicillin plus 2% glucose. DNA was isolated from cultures derived from individual colonies and checked by restriction digestion for orientation. Plasmid DNA containing the correct orientation was sequenced using the Applied Biosystem model 373A automated DNA sequencer to confirm that the fusion was in frame and that no mutations were present.

Four individual clones were used to inoculate cultures of LB-Ampicillin plus glucose and induced with IPTG (1 mM final concentration) at A600 = 0.6. Aliquots were taken at 1, 3 and 5 hours after IPTG addition and protein expression examined by SDS-PAGE. The GST-GCF2 fusion protein was obtained by batch purification with GST-sepharose according to the manufacturers instructions. The GST-GCF fusion protein prepared from IPTG induced cells was dialyzed against saline and used to inoculate rabbits. Antisera was raised in New Zealand rabbits and tested for their ability to immunoprecipitate GCF2 made *in vitro*.

K. Immunoprecipitation

Immunoprecipitations were performed as described by Beguinot et al. Beguinot, L. et al. (1985) *Proc. Natl. Acad. Sci.*, 82:2774-2778. Translation products

labeled with  $^{35}\text{S}$ -methionine were incubated with antisera in the presence of RIPA buffer. The antigen-antibody complexes were washed, dissociated and analyzed by SDS-PAGE.

L. Preparation of cell lysates and cell fractionation

Cell lysates and cell fractionation were performed according to R.B. Dyer and N.K. Herzog (1995) *BioTechniques* 19:192-195. Whole cell lysates and fractionated lysates were subjected to SDS-PAGE and GCF2 presence determined by western blot analysis using a Vectastain ABC kit from Vector Laboratories (Burlingame, CA) and GCF2 antisera.

M. Chromosomal Localization

To localize the chromosomal locus encoding GCF2 gene, a cDNA probe (3.6 kb) was labeled by nick translation with biotin-11-dUTP and used for fluorescence *in situ* hybridization (FISH) as described by Pinkel et al. Pinkel, D. et al. (1986) *Proc. Natl. Acad. Sci. USA* 83:2934-2938. Chromosomes obtained from methotrexate synchronized normal peripheral leukocyte cultures, were penetrated with RNase, denaturated for 2 minutes at 70°C in 2x SSC, 70% (v/v) formamide and hybridized with the DNA probe (200 ng) in 3x SSC, 50% (v/v) formamide, 10% (w/v) dextran sulphate, 2x Denhart's solution, 1% Tween 20 (v/v) and 50 mg human Cot-1 DNA (BRL) probe for 18 hours at 37°C. Posthybridization washing was in 50% formamide-2x SSC at 42°C (3x6 minutes each) and in 0.1xSSC at 60°C (3x6 minutes each). Biotin-labeled DNA was detected by fluorescein isothiocyanate (FITC)-conjugated avidin DCS and antiavidin antibodies (Vector Laboratories). Chromosomes were counterstained with propidiumiodide and examined with a Olympus BH2 epifluorescence microscope.

For each chromosomal spread two consecutive epifluorescent images (FITC and propidium iodide) were recorded by intensified CCD camera connected via image processor (XC-77/C2400, Argus-10, respectively, Hamamatsu Photonics K.K.) to an Apple Macintosh II computer equipped with digitizing board (QuickCapture, Data Translation, Inc.) controlled by NIH's Image. Corresponding 8-bit gray scale digital images were enhanced for sharpness and contrast with NIH's Image and Microfrontier's Enhance, precisely overlaid (GeneJoin Layers developed by T. Rand and S. G. Ballard, Yale University) and the merged by related GeneJoin MaxPix software. Selected merged images were adjusted for size, pseudocolored using interactive graphic package

(PixelPaint Professional, SuperMac Technology) and digitally printed on Textronix's Phaser IISDX dye sublimation color printer (Tektronix Corporation). To obtain chromosome banding, the coverslips from the slides with recorded labeled metaphases were removed in 100% ethanol bath for 10 min, air-dried, incubated in wash buffer (4xSSC-0.1% Tween 20), stained with DAPI (0.2 mg/ml) and mounted in antifade (pH = 8-12). If the resolution of banding after DAPI staining was not satisfactory, the slides were destained in three washes in wash buffer followed by ethanol series (60, 90 and 100%), treated for 30 seconds with trypsin (GIBCO), diluted in Han's balanced salt solution (1:50) and stained with Wright's stain. N.C. Popescu et al. (1985) *Cytogenet. Cell Genet.* 39:73-74. The chromosome spreads were relocated, recorded, photographed and the images compared.

## II. RESULTS

### A. Differential Hybridization Of GCF1 cDNA Fragments

GCF1 cDNA hybridizes to three mRNA species of 4.5, 3.0 and 1.2 kb in several cell lines. Various fragments of the cDNA hybridized differently to the three mRNA species (Johnson, A.C. et al. "Expression and chromosomal localization of the gene for the human transcriptional repressor GCF." *J. Biol. Chem.* 267:1689-1694). A fragment containing nucleotides 1 to 282 and a fragment containing nucleotides 314 to 961 (SEQ ID NO:4) were prepared by PCR, labeled with <sup>32</sup>P and used in Northern blot hybridization analysis. The fragments were designed so they did not contain the stretch of 21 A residues in the GCF1 cDNA which would hybridize to many RNAs. The fragment containing nucleotides 1 to 282 (SEQ ID NO:4) hybridized to an mRNA of approximately 4.5 kb with virtually no hybridization to other mRNAs (Fig. 4A). In contrast, a fragment containing nucleotides 314 to 961 (SEQ ID NO:4) hybridized very strongly to mRNAs of 3.0 and 1.2 kb but only slightly to the 4.5 kb mRNA. This was true using RNA from both A431 and KB epidermoid carcinoma cell line (OVCAR-3) and a T-cell lymphoma cell line (HUT-102).

### B. GCF2 cDNA Isolation

To isolate the cDNA corresponding to the 4.5 kb mRNA, a cDNA library was prepared from ovarian carcinoma cell mRNA (OVCAR3) and it was screened using the fragment containing nucleotides 1 to 282 (SEQ ID NO:4) as probe. Fourteen

positive clones were isolated and sequenced. The two largest clones, O (1.4 kbp) and Q (2.6 kbp) were sequenced and contained all the sequences of the fourteen clones (Fig. 3). The O clone was sequenced and an open reading frame was detected that extended to the 5' end of the O clone. To obtain additional sequence present at the 5' end of the cDNA, rapid amplification of cDNA ends (RACE) was performed. The end of the open reading frame was obtained with an additional 126 bp 5' untranslated region. The cloned cDNA consists of 3523 bp with an open reading frame of 2256 nucleotides. The GCF2 cDNA has a region of sequence homology with the GCF1 cDNA of 309 bp (98% identity) (Fig. 3). The remainder of the sequence has no further significant homology to GCF1 or any other sequence found in GenBank. The deduced protein sequence of GCF2 is shown in Fig. 2. The amino acid sequence indicates the presence of potential phosphorylation sites for cAMP dependent kinase, calcium dependent kinase and tyrosine kinase. Also, the presence of an N-linked glycosylation site and a nuclear localization sequence are predicted.

#### C. GCF2 mRNA Characterization

To determine the size of mRNA to which the GCF2 cDNA hybridizes, northern blot hybridization analysis was performed. Poly (A)<sup>+</sup> RNA isolated from A431, KB and HUT102 cells was transferred to nitrocellulose and probed with a radiolabeled GCF2 cDNA probe. As compared to an RNA size ladder, a 4.2 kb mRNA hybridized to the GCF2 cDNA (Fig. 6). If comparison is made to ribosomal RNA migration, the size would be 4.5 kb which was the original size estimate. The 4.2 kb GCF2 mRNA was detected in all three cell lines with the highest level found in HUT102 cells.

#### D. Production Of GCF2 In Reticulocyte Lysates And *E. coli*

The open reading frame of the GCF2 consists of 2256 residues and is deduced to encode a protein of 83 kDa. The open reading frame was cloned into the pCITE2A vector and coupled *in vitro* transcription/translation performed in the presence of radiolabeled methionine. The radiolabeled translation product was analyzed on an SDS-polyacrylamide gel. GCF2 made *in vitro* in reticulocyte lysates migrates as a protein of 160 kDa, approximately twice the expected size (Fig. 7). The GCF2 open reading frame was also subcloned into a bacterial expression vector containing a His-Tag

sequence. The protein was expressed in bacteria and the His-Tag protein purified on Ni-NTA resin. The purified GCF2-His-Tag protein was analyzed by SDS-polyacrylamide gel electrophoresis and found to migrate the same as the protein made in reticulocyte lysates (Fig. 8). GCF2 has a pH of 4.4 and contains 22% acidic residues.

The GCF2 deduced protein sequence contains a DNA binding and nuclear localization motif similar to GCF1 (Fig. 5). The GCF2 protein expressed from the open reading frame migrates as a 160 kDa protein on SDS polyacrylamide gels. However, the calculated molecular mass is 83 kilodaltons. This could be due to the acidic nature of the protein (pH = 4.4 and 22% acidic residues) or to an unusual ability to form very stable dimers.

#### E. DNA Binding Studies

The homology between the GCF2 cDNA and GCF1 cDNA is confined to the DNA binding region of GCF1. Gel electrophoretic mobility shift assays were used to determine if GCF2 could also bind DNA. Three EGFR promoter fragments were end-labeled and incubated with GCF2-His. Two fragments, (-384 to -167) and (-105 to -16) bound GCF2 and exhibited altered mobility during polyacrylamide gel electrophoresis (SEQ ID NO:6) (Fig. 9). An EGFR promoter fragment containing residues -167 to -105 (SEQ ID NO:6) did not bind GCF2.

To locate the site(s) of GCF2 binding, DNase I footprinting experiments were performed. GCF2 was shown to bind to one site in the EGFR promoter located between -249 to -233 (SEQ ID NO:6) (Fig. 10). There was no footprint detected between -105 and -16 (SEQ ID NO:6). These results suggests that GCF2 binds with different affinities to different sites. It is also evident that there is weaker binding of GCF2 to the -105 to -16 (SEQ ID NO:6) fragment.

Production of protein from deletion mutants in reticulocyte lysates revealed that the altered migration during SDS-PAGE is associated with the protein sequence between residues 490 and 530. This region includes the putative DNA-binding region and the nuclear localization signal. It contains a sequence stretch of residues where 11 out of 14 are lysine. Charge interactions between this region and acidic regions may result in a protein conformation that has an aberrant migration on SDS polyacrylamide gels.



F. Cotransfection Experiments With Promoter-CAT Constructs

The binding of GCF2 to EGFR promoter fragment indicates a possible effect of GCF2 on EGFR gene activity. Cotransfection experiments were performed to examine the effect of GCF2 on EGFR gene expression. GCF2 cDNA (pCMVGCF2) or the control (pCMVGCFR), in which the cDNA is in the reverse orientation was cotransfected with receptor plasmids containing the chloramphenicol acetyltransferase (CAT) gene under control of either the EGFR promoter (pERCAT6), the SV40 early promoter (pSV2CAT) or the Rous Sarcoma Virus LTR promoter (RSVCAT). As shown in Fig. 11, cotransfection with the GCF2 expression plasmid resulted in significant repression of the expression of all three promoters. The control expression plasmid, pCMVGCF2R has no effect on expression from any of these plasmids. The extent of repression by GCF2 was similar for all three reporter plasmids, 3-4 fold at a 5:1 GCF2/CAT.

GCF2 may be acting as either an active repressor or as a passive repressor. In either case it appears to be a general transcription factor. The binding site determined by DNase I footprinting overlaps a GCF1 binding site and a potential AP2 binding site.

G. Tissue-specific Expression Of GCF2

An antisera has been developed against a bacterially expressed GCF2 fusion protein and used that antisera to examine expression of GCF2 in cultured cell lines. The antisera is reactive against GCF2 expressed using a coupled *in vitro* transcription/translation system (Figure 15).

Western blot analysis of lysates from cultured cell lines revealed that GCF2 migrates as a 160 kd protein (Figure 16). This confirms that the protein size of GCF2 produced *in vitro* is accurate. GCF2 is expressed at a very high level in lysates from Raji cells (Burkitt's lymphoma) and HUT102 cells (T-cell lymphoma) and at least three forms are found. Low level expression was observed in other cell lines and no cross-reaction to a comparable size protein in lysates from mouse cells was detected. Fractionation of HUT102 cells into nuclear and cytosolic extracts and analysis of GCF2 localization resulted in finding GCF2 in both compartments but predominately in the nuclear fraction (Figure 17). Thus, GCF2 appears to be a nuclear protein with post-transcriptionally modified forms.

The expression of mRNA that hybridizes to this clone was analyzed by northern analysis using poly(A)+ RNA blots. An RNA species of approximately 4.2 kb was detected by hybridization and comparison to RNA size markers (Figure 18). The RNA was present in all tissues examined with barely detectable levels in brain and testis. The highest level of was found in peripheral blood leukocytes (PBLs). An alternative size of mRNA (2.9 kb) was highly expressed in skeletal muscle, at levels approximately 15-fold greater than most tissues. This species was also found at low levels in heart tissue along with the 4.2 kb mRNA. In PBLs, a larger 6.6 kb mRNA was evident. In most tissues, a weaker hybridizing mRNA of 2.4 kb was also detected.

In cancer cell lines, the 4.2 kb species predominates and only a barely detectable amount of the 2.4 kb mRNA was found (Figure 19). The expression level varied with the highest level found in a Burkitt's lymphoma cell line (Raji). In breast cancer cell lines, BT-20 and BT-474 express very high levels of GCF2 mRNA (Figure 20). A high level of expression is also detected in ZR-75-1 with much lower levels found in other breast cancer cell lines. Again, the 4.2 kilobase mRNA is detected but not the 2.4 kilobase mRNA.

#### H. Localization Of GCF2 To The Long Arm 20q13.3

The GCF2 gene was localized on normal human chromosomes hybridized with a biotinylated cDNA probe (3.6 kb). In chromosome spreads with low non-specific FITC background hybridization signal consisting of symmetrical fluorescent doublets on sister chromatids was visible in 63 (31.50%) of 200 metaphases-examined. However, using intensifying CCD camera and two consecutive recordings through appropriate filters signal was detected in 142 metaphases (71.00%) at the telomeric region of one or two small submetacentric chromosomes.

After chromosome G-banding was obtained (N.C. Popescu et al., *supra*), the chromosome exhibiting fluorescent doublets was identified as chromosome 20 (Figure 21) and by on-screen analysis of digital images of both labeled and banded chromosomes from 30 metaphases with minimal chromosome overlapping, the signal was assigned at terminal region of the long arm 20q13.3.

The present invention provides novel polynucleotides, polypeptides and methods for their use. While specific examples have been provided, the above

description is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims  
5 along with their full scope of equivalents.

All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted. This includes priority United States Provisional Application 60/016,465, filed April 29, 1996.

WHAT IS CLAIMED IS:

- 1                   1.     A purified GCF2 protein whose amino acid sequence is  
2     substantially identical to the amino acid sequence of SEQ ID NO:2.
- 1                   2.     The purified GCF2 protein of claim 1 which is native GCF2,  
2     whose amino acid sequence is SEQ ID NO:2.
- 1                   3.     The purified GCF2 protein of claim 1 which is a human allelic  
2     variant or an animal cognate of native GCF2 that can be encoded by a polynucleotide  
3     that hybridizes under stringent conditions to the nucleotide sequence encoding native  
4     GCF2 of SEQ ID NO:1 and that is isolatable from a human or animal cDNA or genomic  
5     library.
- 1                   4.     A GCF2 protein analog whose amino acid sequence is not naturally  
2     occurring and which comprises a contiguous sequence of at least 10 amino acids from the  
3     amino acid sequence of native GCF2 (SEQ ID NO:2).
- 1                   5.     The GCF2 protein analog of claim 4 which, when presented as an  
2     immunogen, elicits the production of an antibody which specifically binds to native  
3     GCF2 protein.
- 1                   6.     The GCF2 protein analog of claim 4 that inhibits the expression of  
2     a nucleotide sequence operably linked to the EGFR gene promoter, the RSV promoter or  
3     the SV40 promoter.
- 1                   7.     The GCF2 analog of claim 6 that comprises an amino acid  
2     sequence substantially identical to amino acids 511 to 524 of SEQ ID NO:2.
- 1                   8.     The GCF2 protein analog of claim 6 comprising a contiguous  
2     sequence of 550 acids having substantial sequence identity with an amino acid sequence,  
3     either contiguous or non-contiguous, from native GCF2.

1                   9.     The GCF2 protein analog of claim 6 whose amino acid sequence is  
2 a fragment of the sequence of SEQ ID NO:2.

1                   10.    The GCF2 protein analog of claim 6 further comprising a  
2 polyhistidine tag.

1                   11.    A recombinant polynucleotide comprising a nucleotide sequence of  
2 at least 25 contiguous nucleotides from nucleotides 128 to 1384 or 1694 to 2310 of SEQ  
3 ID NO:1.

1                   12.    A recombinant polynucleotide comprising a nucleotide sequence  
2 that codes for the expression of a polypeptide whose amino acid sequence comprises a  
3 contiguous sequence of at least 10 amino acids selected from amino acids 1 to 752 of  
4 SEQ ID NO:2.

1                   13.    The recombinant polynucleotide of claim 12 whose nucleotide  
2 sequence is substantially identical or identical to the nucleotide sequence of SEQ ID  
3 NO:1.

1                   14.    The recombinant polynucleotide of claim 13 whose nucleotide  
2 sequence is identical to the nucleotide sequence of SEQ ID NO:1.

1                   15.    The recombinant polynucleotide of claim 12 wherein the nucleotide  
2 sequence that encodes the at least 10 amino acids is a nucleotide sequence from SEQ ID  
3 NO:1.

1                   16.    The recombinant polynucleotide of claim 15 wherein the nucleotide  
2 sequence codes for the expression of a protein which, when presented as an immunogen,  
3 elicits the production of an antibody which specifically binds to native GCF2 protein.

1                   17.    The recombinant polynucleotide of claim 16 wherein the nucleotide  
2 sequence codes for the expression of a protein that inhibits the expression of a nucleotide  
3 sequence operably linked to the EGFR gene promoter.

1                   18.    The recombinant polynucleotide of claim 16 wherein the nucleotide  
2                   sequence comprises a sequence coding for an amino acid sequence substantially identical  
3                   to amino acids 511 to 524 of SEQ ID NO:2.

1                   19.    The recombinant polynucleotide of claim 12 wherein the expression  
2                   control sequences are eukaryotic expression control sequences.

1                   20.    A recombinant host cell transfected with an expression vector  
2                   comprising expression control sequences operatively linked to a nucleotide sequence that  
3                   codes for the expression of a polypeptide whose amino acid sequence comprises a  
4                   contiguous sequence of at least 10 amino acids selected from amino acids 1 to 752 of  
5                   SEQ ID NO:2.

1                   21.    An isolated polynucleotide probe comprising at least 15 nucleotides,  
2                   that specifically hybridizes with a unique nucleotide sequence of nucleotide sequence of  
3                   SEQ ID NO:1 or with its complement.

1                   22.    The isolated polynucleotide probe of claim 21 comprising at least  
2                   25 nucleotides.

1                   23.    The isolated polynucleotide probe of claim 21 further comprising a  
2                   label.

1                   24.    An antibody that specifically binds native GCF2 protein.

1                   25.    The antibody of claim 24 that is a polyclonal antibody.

1                   26.    The antibody of claim 24 that is a monoclonal antibody.

1                   27.    The antibody of claim 24 that is a humanized antibody.

1                   28.    The antibody of claim 24 that is an antibody fragment.

1                   29.    A method for detecting GCF2 protein in a sample comprising  
2    contacting the sample with a GCF2 binding substrate, and detecting the presence of a  
3    GCF2 protein/substrate bound complex, whereby the presence of the complex indicates  
4    the presence of GCF2 protein in the sample.

1                   30.    The method of claim 29 wherein the sample is a nuclear extract  
2    from a cell.

1                   31.    The method of claim 30 wherein the cell is a breast cell, an ovary  
2    cell, a cervix cell or a kidney cell.

1                   32.    The method of claim 29 wherein the GCF2 binding substrate is a  
2    polynucleotide comprising the nucleotide sequence of SEQ ID NO:3 from the EGFR  
3    promoter.

1                   33.    The method of claim 29 wherein presence of the complex is  
2    detected by determining a mass of the complex greater than an expected mass of a GCF2  
3    protein whereby a greater mass indicates that GCF2 protein is bound with the binding  
4    substitute.

1                   34.    The method of claim 29 wherein the presence of the complex is  
2    detected by immunoassay.

1                   35.    The method of claim 29 for determining the amount of GCF2  
2    binding activity in a sample wherein detecting the presence of complex comprises  
3    determining the amount of complex, the method further comprising comparing the  
4    amount with a standard amount of complex based on known amounts of native GCF2  
5    protein.

1                   36.    A kit comprising a GCF2 binding substrate and an anti-GCF2  
2    antibody.

1                   37.    The kit of claim of claim 36 wherein the substrate of the antibody  
2    is labeled.

1                   38.    The kit of claim 36 further comprising a GCF2 protein or GCF2  
2    protein analog having DNA binding activity, useful as a standard.

1                   39.    A kit comprising a GCF2 binding substrate and a GCF2 protein or  
2    GCF2 protein analog having DNA binding activity.

1                   40.    The kit of claim of claim 39 wherein the substrate or the GCF2  
2    protein or GCF2 protein analog is labeled.

1                   41.    A method for isolating DNA sequences that bind to a GCF2 protein  
2    comprising contacting a GCF2 protein or a GCF2 protein analog that has GCF2 binding  
3    activity with a DNA library.

1                   42.    A method for inhibiting in a cell the activity of a promoter  
2    regulated by GCF2 comprising providing a GCF2 protein or an active GCF2 protein  
3    analog to the cell.

1                   43.    The method of claim 42 wherein the GCF2 protein or the active  
2    GCF2 protein analog is provided to the cell by transfecting the cell with a recombinant  
3    polynucleotide comprising expression control sequences operatively linked to a nucleotide  
4    sequence coding for the expression of a polypeptide whose amino acid sequence  
5    comprises a contiguous sequence of at least 10 amino acids selected from amino acids 1  
6    to 752 of SEQ ID NO:2.

1                   44.    The method of claim 43 wherein the promoter is the EGFR  
2    promoter, the RSV promoter or the SV40 promoter.

1                   45.    The method of claim 44 wherein the promoter is the EGFR  
2    promoter operably linked to a nucleotide sequence that codes for the expression of  
3    EGFR.



1                   46.    The method of claim 45 wherein the promoter EGFR sequence is  
2   endogenous to the cell.

1                   47.    A method useful for restoring GCF2 binding activity in a cancer  
2   cell that exhibits reduced GCF2 binding activity comprising transfecting the cancer cell  
3   with a recombinant polynucleotide comprising expression control sequences operatively  
4   linked to a nucleotide sequence coding for the expression of a polypeptide whose amino  
5   acid sequence comprises a contiguous sequence of at least 10 amino acids from the  
6   sequence of native GCF2 (SEQ ID NO:2), wherein at least 10 amino acids of contiguous  
7   sequence are selected from amino acids 1 to 752 of SEQ ID NO:2 and wherein the  
8   protein inhibits the expression of a nucleotide sequence operably linked to the EGFR  
9   gene promoter.

1                   48.    A therapeutic method for inhibiting the growth of a cancer cell that  
2   over-expresses the EGF receptor in a subject comprising transfecting the cancer cells  
3   with a recombinant polynucleotide comprising expression control sequences operatively  
4   linked to a nucleotide sequence coding for the expression of a polypeptide whose amino  
5   acid sequence comprises a contiguous sequence of at least 10 amino acids selected from  
6   amino acids 1 to 752 of SEQ ID NO:2 and wherein the protein inhibits the expression of  
7   a nucleotide sequence operably linked to the EGFR gene promoter.

1                   49.    A therapeutic method for inhibiting the growth of cancer cells that  
2   over-express the EGF receptor in a subject comprising administering to the subject an  
3   effective amount of a GCF2 protein or active GCF2 protein analog.

1                   50.    A method of producing a GCF2 protein or GCF2 protein analog  
2   comprising transfecting a host cell transfected with an expression vector comprising  
3   expression control sequences operatively linked to a nucleotide sequence that codes for  
4   the expression of a polypeptide whose amino acid sequence comprises a contiguous  
5   sequence of at least 10 amino acids selected from amino acids 1 to 752 of SEQ ID NO:2  
6   and culturing the cell to express the GCF2 protein or GCF2 protein analog.

1           51. A method of detecting GCF2 mRNA or cDNA in a sample  
2 comprising the steps of: (a) contacting the sample with a polynucleotide probe or primer  
3 that specifically hybridizes to GCF2 nucleotide sequences; and (b) detecting specific  
4 hybridization of the probe to GCF2 whereby specific hybridization provides a detection  
5 of GCF2 mRNA or cDNA in the sample.

1           52. The method of claim 51 wherein the step of detecting comprises  
2 amplifying the GCF2 nucleotide sequences by RT-PCR.

1           53. The method of claim 51 wherein the sample is a cell sample and the  
2 step of contacting comprises *in situ* hybridization with a labeled probe.

1           54. A method for aiding in the diagnosis of cancer comprising the steps  
2 of: (a) determining a diagnostic value by detecting one or more GCF2 mRNA species in  
3 a subject sample; and (b) comparing the diagnostic value with a normal range of the  
4 species in a control cell sample whereby a diagnostic value that is above the normal  
5 range is a positive sign in the diagnosis of cancer.

1           55. The method of claim 54 wherein the mRNA species are about 4.2 kb  
2 and about 2.4 kb and diagnostic values of the 4.2 kb species above the normal range or  
3 values of the 2.4 kb species below the normal range provides a positive sign in the  
4 diagnosis of cancer.

1           56. The method of claim 54 wherein the cancer is breast cancer, a B-cell  
2 lymphoma or a T-cell lymphoma.

1           57. A method of detecting a chromosomal translocation of a GCF2 gene  
2 comprising the steps of (a) hybridizing a labeled polynucleotide probe of claim 23 to a  
3 chromosome spread from a cell sample to determine the pattern of hybridization and (b)  
4 determining whether the pattern of hybridization differs from a normal pattern.

1           58. A method of detecting polymorphic forms of GCF2 comprising  
2 comparing the identity of a nucleotide or amino acid at a selected position from the  
3 sequence of a test GCF2 gene or polypeptide with identity of the nucleotide or amino  
4 acid at the corresponding position of native GCF2, whereby a difference in identity  
5 indicates that the test polynucleotide is a polymorphic form of GCF2.

## NUCLEOTIDE SEQUENCE OF THE GCF2 cDNA

GACTTAGCGGCGAGTGGCTCCGTCTCCGCGGACAGAGCGCGCGCCCCCTGGCCCGGCCCGGAGGGGCTC 70  
 CCGGCGCGGTCGCCGAGCATTTCCCGCGGGTGGAGCGGGCCGAGCCAGGCAGGATGACCAGCCCCGCGG 140  
 CCGCTCAAAGCCGGGAGATCGACTGTTTGGAGCTTGAAGCGCAGAAGCTGGCGGAAGCCTGGCTCGCTGC 210  
 AAAACGGGCGGCGCGCGGAGGCTCGCGAGATCCGCATGAAGGAGCTGGAGCGGCAGCAGAAGGAGGTA 280  
 GAAGAGAGACCAGAAAAAGATTTTACTGAGAAGGGGTCTCGTAACATGCCGGGCTGTCTGCAGCCACGC 350  
 TGGCCTCTCTGGGTGGGACTTCCTCTCGGAGAGGCAGCGGAGACACCTCCATCTCCATCGACACCGAGGC 420  
 ATCCATCAGGGAATCAAGGACTCTCTAGCAGAAGTTGAAGAGAAATATAAGAAGGCTATGTTGCTGGAGC 490  
 GCTCAGCTAGACAATGAAAAGACAACTTCATGTACCAGGTTGATACCCATAAAGATATGTTGCTGGAGC 560  
 TTGAAGAACAGCTGGCTGAATCTAGCGGCGAGTACGAAGAGAAAAACAAGAAATTTGAAAGGGAAAAACA 630  
 CGCCACAGTATACTGCAATTTCACTTTGCTGAAGTCAAGGAGGCCCTGAAGCAAAGAGAGGAAATGCTC 700  
 GAGAAACATGGAATAATCCTAAATTCAGAAATAGCTACCAATGGAGAGACTTCCGACACCCCTCAATAATG 770  
 TTGGATACCAAGGCTCTACCAAGATGACAAAAGAAGAGTTAAATGCCCTCAAGTCGACAGGGGATGGGAC 840  
 CCTAGGAAGAGCCAGTGAAGTGGAGGTGAAAAATGAAATCGTGGCGAATGTGGGGAAAAAGAGAAATCTG 910  
 CACAATACTGAGAAAGAGCAACACGCAGAGGACACAGTGAAGGACTGTGTGGACATAGAGGTATTCCTTG 980  
 CTGGTGAGAATACCGAGGACCAGAAATCCTCTGAAGACACTGCCCCATTCTAGGAACCTTAGCAGGTGC 1050  
 TACCTATGAGGAACAGGTTCAAAGCCAAATTTCTTGAGAGCAGTTCTCTCCCTGAAAACACAGTACAGGTT 1120  
 GAGTCAAATGAGGTGATGGGTGCACCTGATGACAGGACCAGAACTCCCTTGAGCCATCCAATGTTTGA 1190  
 GTGACTTAGATGGTGGAAACACACAGAGAATGTGGGAGAGGCAGCAGTGAAGTGAAGAGCAGGC 1260  
 AGGCACAGTGGCCTCGTCTCTTAGGGCATAGTGTGACACAGTTTATCATGTGACAAATGTATGGTA 1330  
 GAGGTCCCCCAAGAGTTAGAGACAAGCACAGGCGATAGTTTAGAGAAAGAATTCACCAACCAGGAAGCAG 1400  
 CTGAGCCCAAGGAGGTTCCAGCGCACAGTACAGAAGTAGGTAGGATCACAACGAAGAAGAGGGTGAAGA 1470  
 AACAGGATTAAGGGACGAGAAACCAATCAAGACAGAAGTTCCTGGTCTCCAGCAGGAAGTGAAGGCAAC 1540  
 TGTCAAGGAGCGACAGGTCCAAGTACAGTACACTCAAATGAACCTTAGATATGAAAGAGCCCGATG 1610  
 AAGAAAAGAGTGACCAACAGGGGAGAGGCATTGGACTCATCGCAGAAGAAGACAAAGAAAAGAA 1680  
 AAACAAGAAGAAAAATCCCCAGTACCCGTAGAAACCTTTAAAGATGTTAAAAAGAGTTAACGTTATCAG 1750  
 AACACAGATTTAAGTGAATTAAGGAAGAAGAGCAGGTAAAGTCTACTGACAGAAAGTCAGCAGTGGAA 1820  
 CCCAAACGAGGTGACTGAAAATCCAAACAGAAAAATTCAGCAGAAAGCAGTGAAGATGTTGATTGTCC 1890  
 GGAGATCCTAAATTAAGTTGGATGGAAAACTTGACCAAGAAGGTGATGATGTACAACAGCAGCTGAG 1960  
 GAGGTACTAGCTGGAGACACATTAGATTTTGAGGATGACACCGCTCAATCATCAGGCCCGAGGGCTG 2030  
 GTGGTGAAGAATTAGATGAAGGTGTTGCAAAAGATAATGCTAAATAGATGGTGCCACTCAAAGCAGTCC 2100  
 TGCAGAACCAAGAGCGAAGACGCAGATCGCTGCAACCTGCCGAACATGAAAGTCCCTCACAGGACATT 2170  
 AGTGATGCCTGTGAAGCAGAAAGTACAGAGAGGTGTGAGATGTGAGATGTGAGAACATCCAAGTCAGACCGTCAGGA 2240  
 AAGCTTTAGACAGCAATAGCCTAGAGAACGATGACTTGTGCGCACCCAGGAAGAGAGCCAGGGCACTTCAA 2310  
 TCCAGAAAGCAGAGAAGATACCAGAGGAGGGAATGAGAAGGGCAAAAGCAAAGAAGACTGTACCATGTCC 2380  
 TAAGCTGAGGCAGGCGCGGCGGTGCACAGGAAGTCTCAGTGTGAAGGGGTCTTTCTCTCCACTGC 2450  
 CAATGTAAGTAGAATGTTCTAAATTCATAGAGAGGCAGTGTATGACAATTACCAGGTGCTCTACTGCTTT 2520  
 AAGTTATAGACTGTTACTTGTAGATTTCCATGTAATCATTGAGGTTATCACCAGATTAGAAAGACATAT 2590  
 TTGTTATCAGTGTTACGTTCTAATTGAGAGCATTCAGTAGTATCAAAACAATAATGTCTACTGTTTATAGT 2660  
 CCACTTAATAAAAAATAGAGGCATTTACTATTTGCCCTTAGGCTGATAGGAATGTGGGTTTCTTGACCAAA 2730  
 TATATCAGCATCTAATTGAAATGACCAAAATAGCATTCTTAGACTTCTGTATTATGAATATAAATTGATAT 2800  
 TAAATTAATGTCTTGTTCACATATGTGTACTTTTCAATTTGATTTTAAATGTACATTATAACCTGTATG 2870  
 GTATTTTATTTAAAGGAGATAAAACAGCCAAATAGCAAATAGGTCACTGAATGATAAGATTTGCACCTTAG 2940  
 TATGAGTCTTTGATCTTGAACCGATACTTTTGGATCTCATTGTTGATATACCTGAATTTACTTTGTAAGA 3010  
 GATTTTAACCTTCACTTCATGCTGATGATGTATCAAAATCATTATATAGAAAGATTTAAAGTTTTTCTG 3150  
 GAAGTGATATATGTCAAATTACATTTCTACTGCAGTATTTGAGCAGGGACAGTCATTTTAAATGTTT 3220  
 TTGGCCGGGCGTGGTGGCTCATGCCTGTAATCTCAGTACATTTGGGAGGCCAAGGCAGGTGGATCACCTGA 3290  
 GGTCAAGAGTTTCGAGGCCAGCCTGGCCAAACATGGTGAAACCCCTGCTCTACTAAAAATACAAAAATTGG 3360  
 CCGGCGCTGATGGTGGGCGCCTGTAATCCAGCCACTCCAGAGCTGAGGCAGGAGAATCGCTTGAACCT 3430  
 GCGAGGCAGAGATTGCAGTGAGCCAAGATCAAGCCATTGTACTCCAGCCTGGACAACAAGAGCGAAACTC 3500  
 TGTCTAAAAA

Figure 1

MTSPAAAQSREIDCLSLEAQLAEAWLAAKRAAAREIRMKELERQOK 50  
EVERPEREKDFTEKGSRNMPGLSAAATLASLGGTSSRRGSGDTSISIDTEAS 100  
IREIKDSLAEVEEKYKKAMVSNALDNEKTNFMYQVDTLKDMLLEEEQL 150  
AESRRQYEEKNKEFEREKHAHSILQFQFAEVKEALKQREEMLEKHGIILN 200  
SEIATNGETSDTLNNVGYQGPTKMTKEELNALKSTGDTLGRASEVEVKN 250  
EIVANVGKREILHNTEKEQHTEDTVKDCVDIEVFPAGENTEDQKSS~~ETA~~ 300  
PFLGTLAGATYEEQVQSQILESSSLPENTVQVESNEVMGAPDDRT~~RTPLE~~ 350  
PSNCWSDLDGGNHTENVGEAAVTQVEEQAGTVASCPLGHSDDTVYHDDKC 400  
MVEVPQELESTGHSLEKEFTNQEAAPKEVPAHSTEVG~~RDNHEEEGEET~~ 450  
GLRDEKPIKTEVPGSPAGTEGNCQEATG~~PTVDTQNEPLDMKEPDEEKSD~~ 500  
QQGEALDSSQKTKNKKKKKKKKSPVPVETLKDVKKELTYQNTDLSEIK 550  
EEEQVKSTDRKSAVEAQNEVTENPKQKIAAESSENVD~~CPENPKIKLDGKL~~ 600  
DQEGDDVQPAEEVLADGDTLDFEDDDTAQSSGPRAGGEELDEGVAKDNAK 650  
IDGATQSSPAEPKSEAD~~ADRCTLPEHESPSQDISDACEAESTERC~~EMSEHP 700  
SQTVRKALDSNSLEND~~DL~~SAPGREPGHFN~~PESREDTRG~~NEKGKSKEDCT 750  
MS 752

THIS PAGE BLANK (USPTO)

3/20

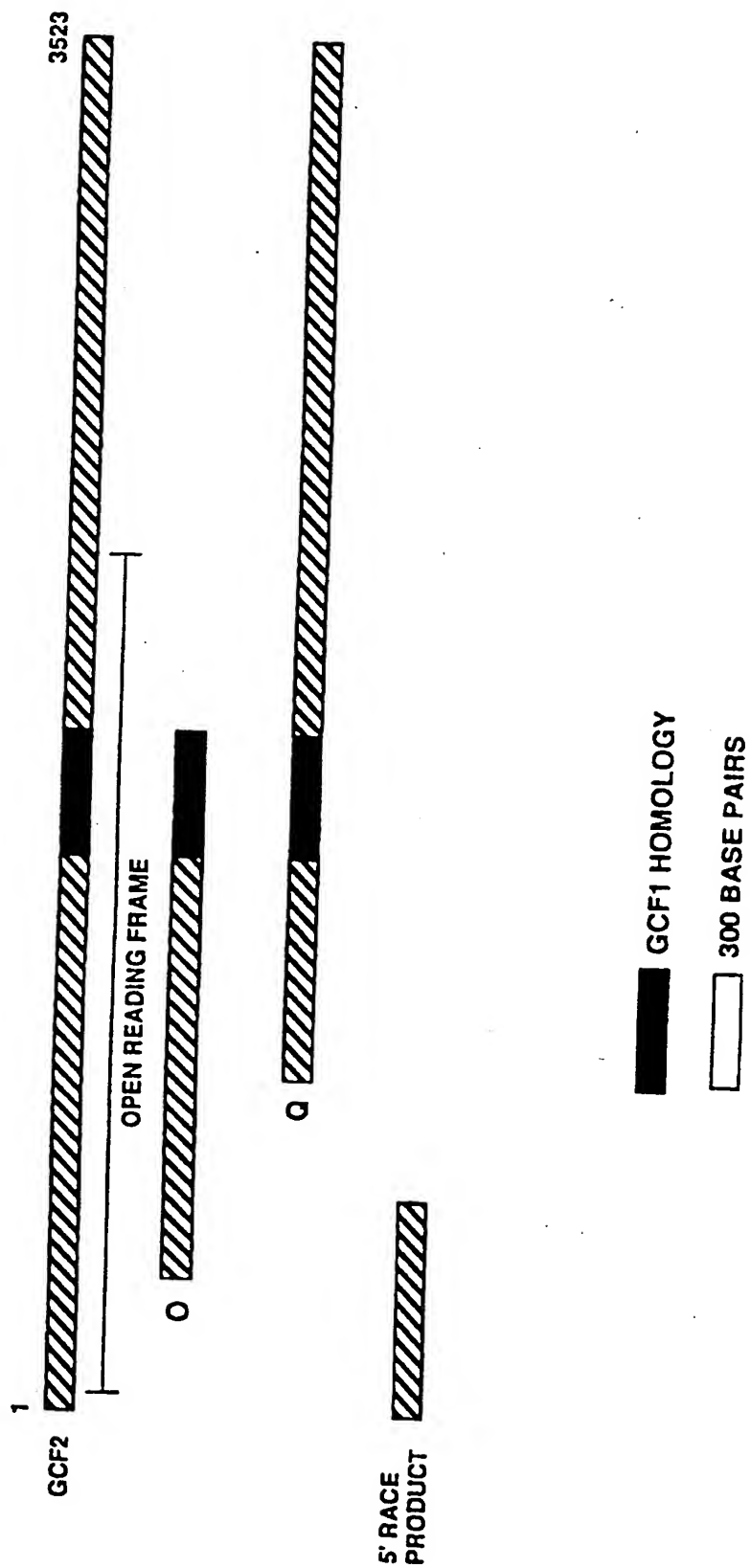


Figure 3

**THIS PAGE BLANK. (USPTO)**



4/20

A. 1 TO 282

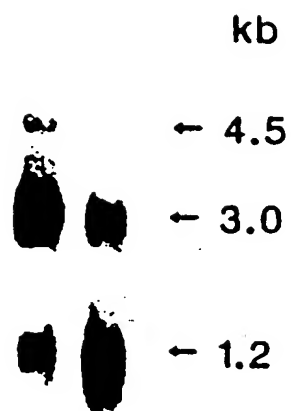
A431  
KB



1 2

B. 314 TO 961

A431  
KB



1 2

Figure 4

**THIS PAGE BLANK (USPTO)**

GCF2: 1385 ACCAACCCAGGAAGCAGCTGAGCCCAAGGAGGTTCCAGCGCACAGTACAGAAAGTAGGTAGG 1444  
|||||  
GCF1: 1 ACCAACCCAGGAAGCAGCTGAGCCCAAGGAGGTTCCAGCGCACAGTACAGAAAGTAGGTAGG 60  
GCF2: 1445 GATCACAAACGAAAGAGAGGGTGAAAGAACAGGATTAAGGGACGAGAAACCAATCAAGACA 1504  
|||||  
GCF1: 61 GATCACAAACGAAAGAGAGGGTGAAAGAACAGGATTAAGGGACGAGAAACCAATCAAGACA 120  
GCF2: 1505 GAAGTTCCCTGGTTCTCCAGCAGGAACTGAGGGCAACTGTCTCAGGAAGCGACAGGTCCAAAGT 1564  
|||||  
GCF1: 121 GAAGTTCCCTGGTTCTCCAGCAGGAACTGAGGGCAACTGTCTCAGGAAGCGACAGGTCCAAAGT 180  
GCF2: 1565 ACAGTAGACACTCAAAATGAACCCCTTAGATATGAAAGAGCCCGATGAAGAAAGAGTGAC 1624  
|||||  
GCF1: 181 ACAGTAGACACTCAAAATGAACCCCTTAGATATGAAAGAGCCCGATGAAGAAAGAGTGAC 240  
GCF2: 1625 CAACAGGGAGAGGCATTGGACTCATCGCAGAGAGACAAAGACAAAGAAAGAAAAAC 1684  
|||||  
GCF1: 241 CAACAGGGAGAGGCATTGGACTCATCGCAGAGAGACAAAGACAAAGAAAGAAAAAC 300  
GCF2: 1685 AAGAAGAAA 1693  
|||  
GCF1: 301 AAAAAAAA 309

Identities = 305/309 (98%), Positives = 305/309 (98%), Strand = Plus

**THIS PAGE BLANK (USPTO)**

6/20

BEST AVAILABLE COPY

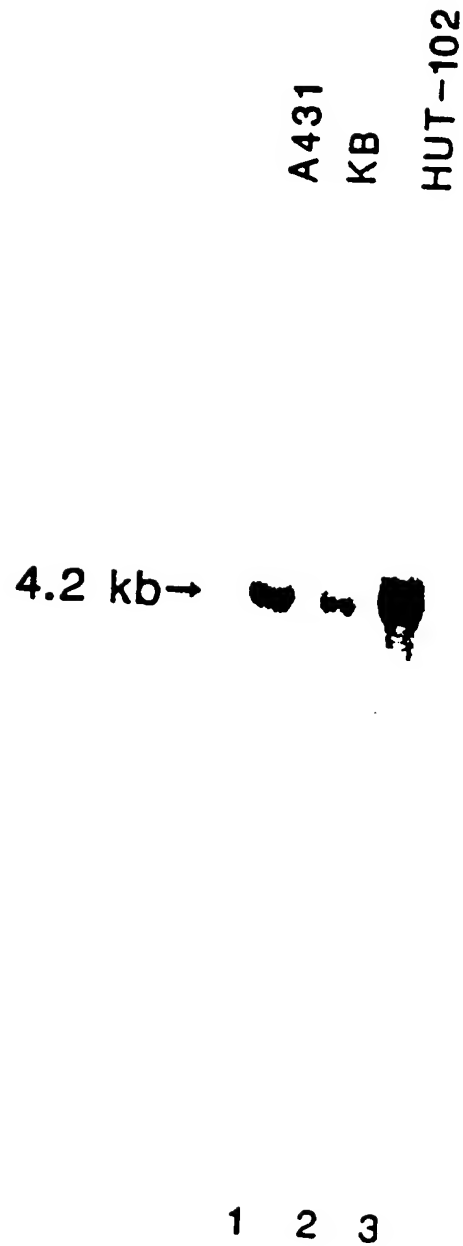


Figure 6

7/20

BEST AVAILABLE COPY

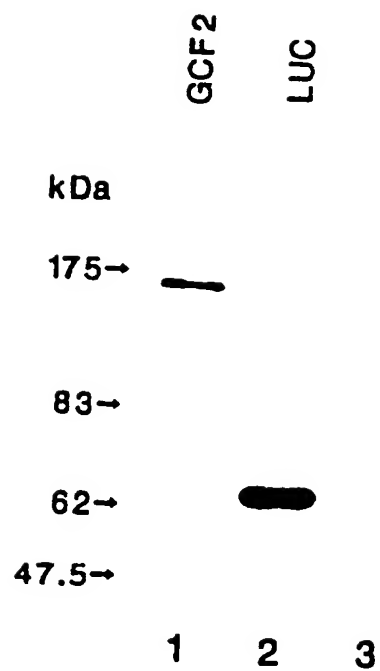


Figure 7

8/20

BEST AVAILABLE COPY

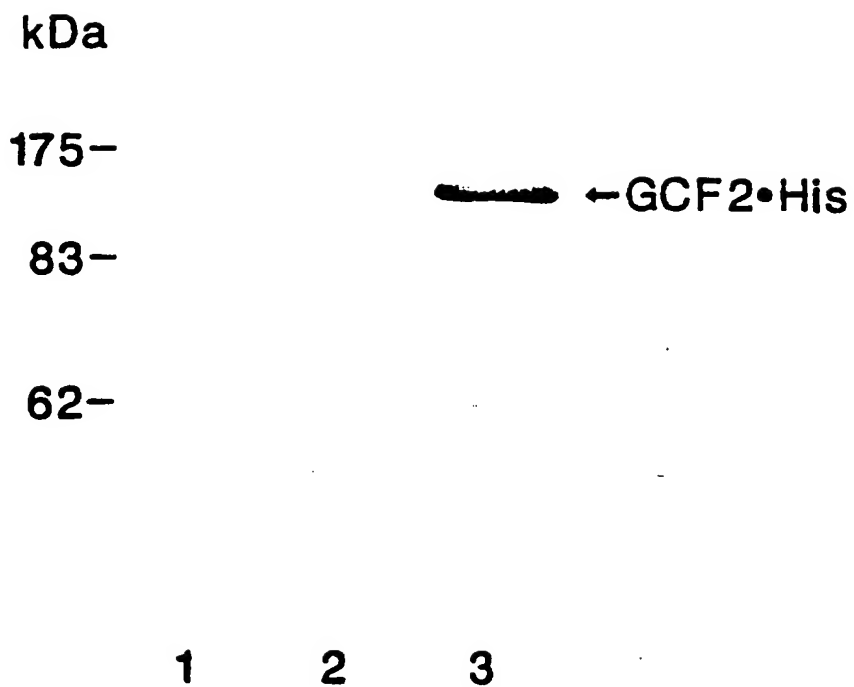


Figure 8

9/20

BEST AVAILABLE COPY

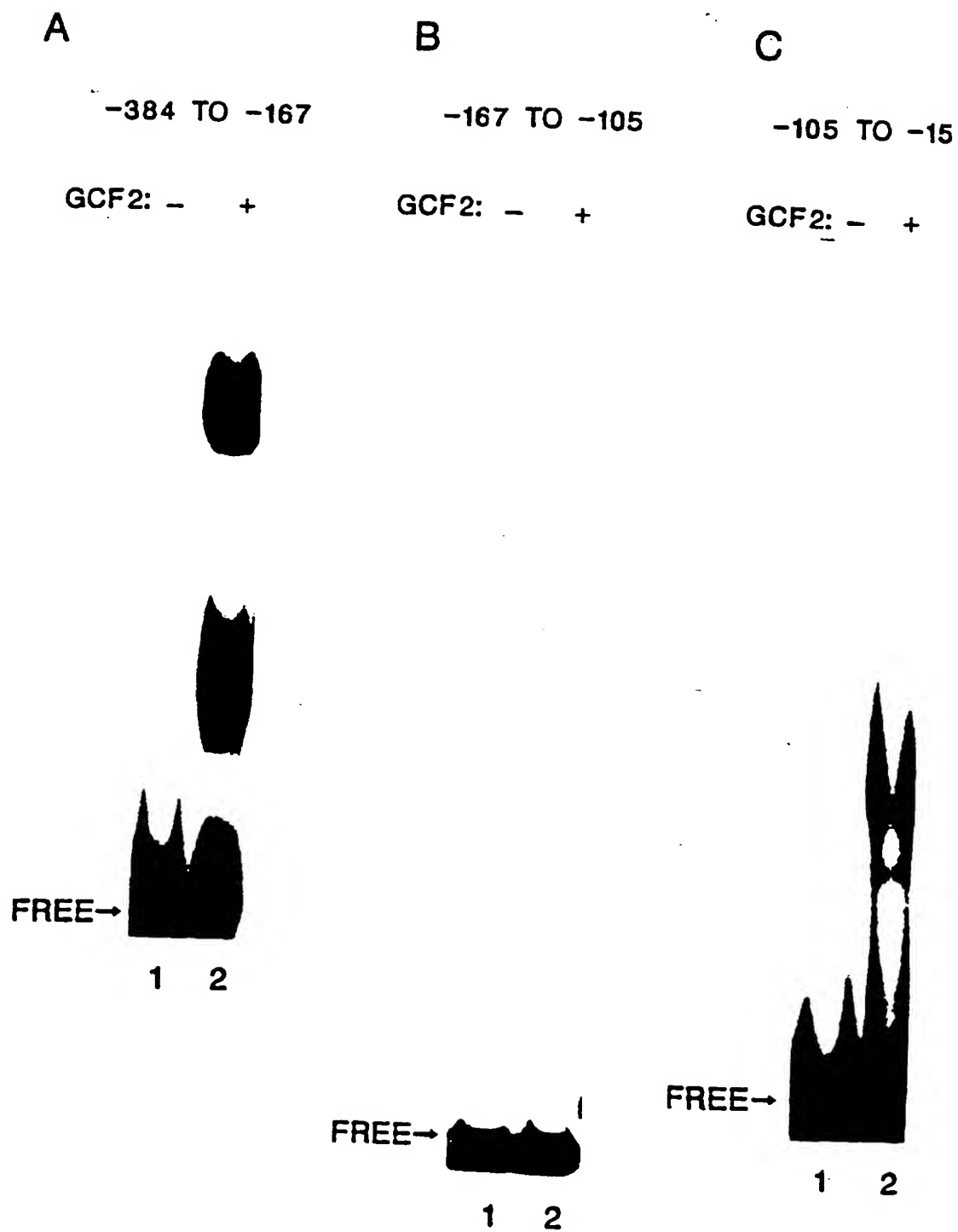
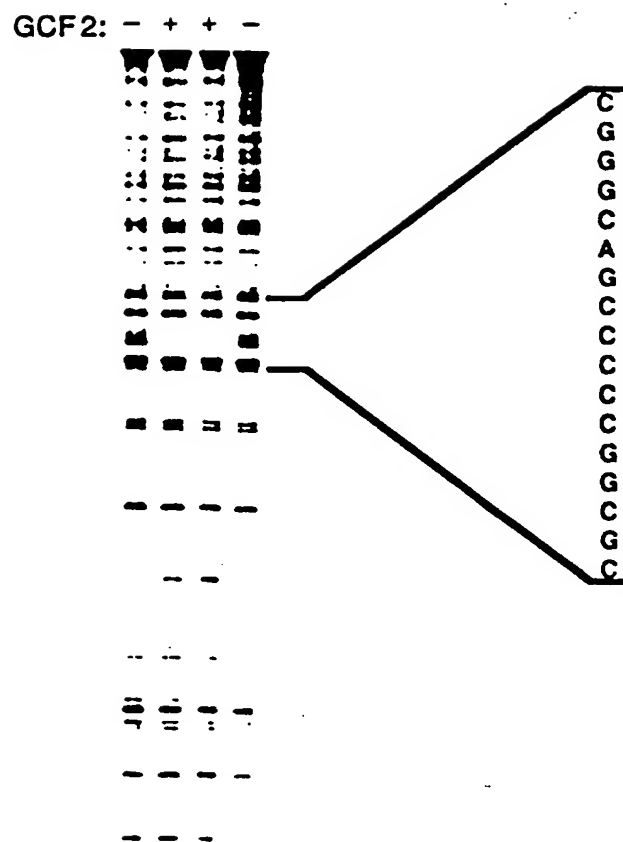


Figure 9



10/20



1 2 3 4

Figure 10

11/20

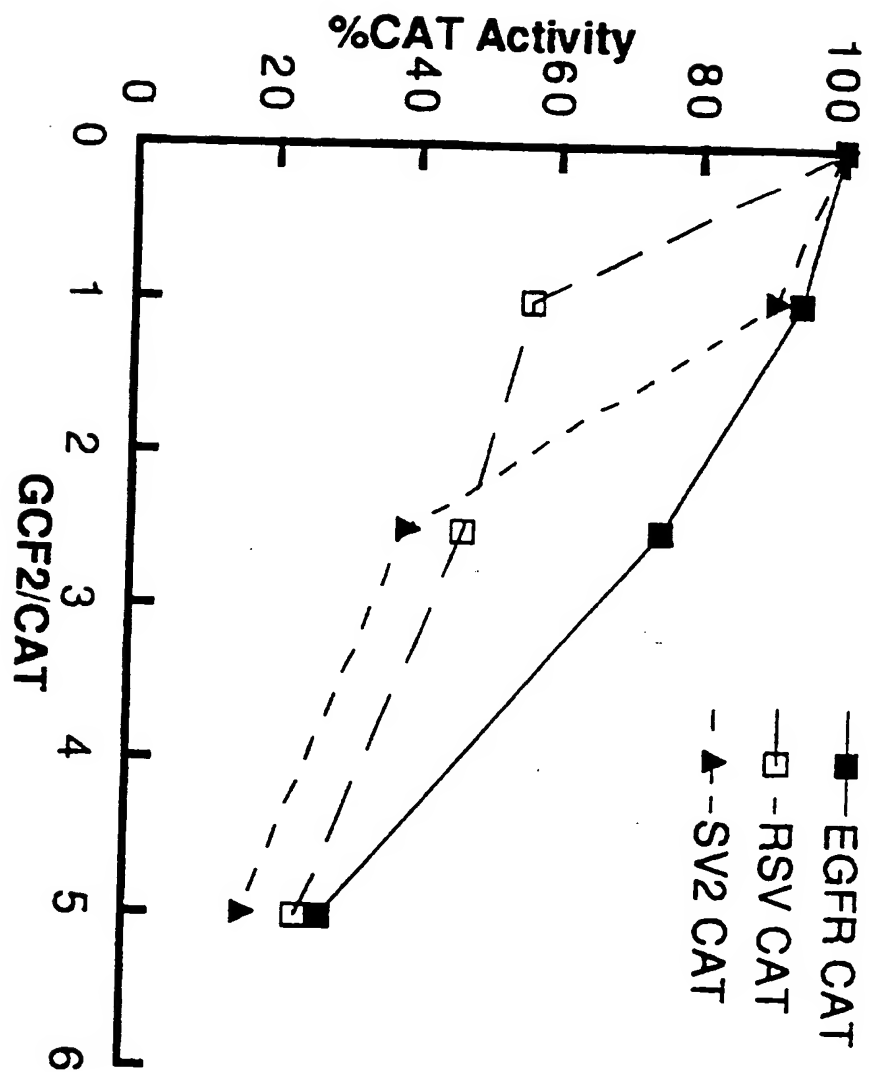


Figure 11

12/20

## NUCLEOTIDE SEQUENCE OF GCF1

ACCAACCAGGAAGCAGCTGAGCCCAAGGAGGTTCAGCGCACAGTACAGAAAGTAGGTAGGGATCACAACG	70
AAGAAGAGGGTGAAGAAACAGGATTAAGGGACGAGAAACCAATCAAGACAGAAGTTCTTGGTTCTCCAGC	140
AGGAACCTGAGGGCAACTGTCAGGAAGCGACAGGTCCAAGTACAGTAGACACTCAAAATGAACCCCTTAGAT	210
ATGAAAGAGCCCGATGAAGAAAAGAGTGACCAACAGGGAGAGGCATTGGACTCATCGCAGAAGAAAGACAA	280
AGAACAAGAAAAAAAAAAAAAAAAAAGCGGGTTCTAGGGCGCGGGCGCTCGGGCCTCGGCCATGGC	350
TCACAGGCCGAAAAGGACTTTTCGGCAGCGCGCGGCTGATTCCAGCGACAGCGATGGCGCCGAGGAGTCG	420
CCTGCTGAGCCTGGGGCGCCGAGGGAACCTTCGGTCCCGGGTTCTGCGGAGGAAGAGCCGCCCTCTGGAG	490
GAGGCCGCGCGCAGGTGGCGGGACTGCCCCACCGGGTTCTGGGGCCCTCGTGGCCGGGGCCGGGTCTGGGC	560
GAGTCCCCGGCTGCCACCAAGCGGCTCCCCCGCGCGGACGAAGGCTCAGAATCCAGAACCCCTTGATGTG	630
TCCACAGATGAAGAGGATAAAATACATCACTCCTCAGAAAGTAAAGGATGATCAGGGTTGTCTCTTGACA	700
GTTCTAGCTCTCTTGCAGAAAAAGAACTTTTCATCAACAGTTAAGATCCAGATGCAGCTTTTATTCAGGC	770
AGCCCGCAGAAAACGTTGAATTGGCCAGGGCCCAAGATGACTATATTTCTTTGGATGTACAACATACCTC	840
CTCCATCTCTGTAAGCAGAAATGAAGAAACAAGTGAAGAAAGTCAGGAAGATGAAAAGCAAGATACTTGG	910
GAACAACAGCAAAATGAGGAAAGCAGTTAAATCATAGAGGAAAGAGACATAGATCTTTCTGTGGCAGTG	980
GATCTTCAAAAGTGAAGAAATTTGATACCTTCCATTTTCATTTCGGCCAGTAAATTTAGAAATTATAAGAA	1050
GCAATTAAATACTAGATTAACATTACTACAGGAACCTCACCGCTCACACCTGAGGGAGTGAAGAAATAC	1120
GTACAAGATGTCAAAAGCTCAAGAGTACCATCCAGAACCTAGAGAGTTCATCAAAATCAAGCTCTAAAT	1190
GTAATTTCTATAAAAGCATGAAAATTTATGTGGAATAATTTAATTGACTGCCTTAATGAAAAGATTATCAA	1260
CATCCAAGAAATAGAAATCATCCATGCAATGCATTCCTTTTAAACAAGCTATGACCTTTATGAAACGCAGG	1330
CAAGATGAATTAACATGAATCAACGTATTACCAACAGTTATCACGCAAGATGAGACATCCACAAGTG	1400
GAACTTCTCAGTAGATGAAAAAATCAGTGGATTTTAGAAGAGATTGAATCTCGAAGGACAAAAAGAAG	1470
ACAAGCAAGGGTGCTTTCTGGGAATTGTAACCATCAGGAAGGAACATCTAGTGATGATGAAGCTGCCTCA	1540
GCAGAGATGATTGACTTCCAAAAAAGCCAAGGTGACATTTTACAGAAACAGAAAGATTTTGAAGAAG	1610
TGCAAGATGATTTTGTAAACATCCAGAAATATTTGTTGAAATTTTCAGCAATGGCGAGAAAAGTTTCTGA	1680
CTCCTATTATGAAGCTTTTATTAGTTTATGCATACCAAGCTTTTAAATCCCCTAATACGAGTTTCAGTTG	1750
ATTGATTGGAATCCTCTTAAGTTGGAATCCACAGGTTTAAAGAGATGCCATGGTTCAAATCTGTAGAAG	1820
AATTTATGGATAGCAGTGTAGAAGATTCAAAGAAGGAAAGTAGTTTCAGATAAAAAAGTCTTGCTGCAAT	1890
CATCAACAAAACAATTATTTCCCGACTTACAGACTTTGTAGAATTCCTTTGGGATCCTTTGTCAACCTCA	1960
CAGACAACAAGTTTAATAACACATTGCAGAGTGATTCTTGAAGAATTCCTTGTGAAAATGAAGTTA	2030
GTAAGGCAGACAGGATTACTTAAATCCATTGTTTCAAGAATGAAAAAGGCAGTAGAAGATGATGTTTT	2100
TATTCCTCTGTATCCAAAGAGTGCTGTAGAAAAACAAACATCACCTCATTCAAAGTTCCAAGAAAGACAG	2170
TTCTGGTCAGGCCTAAAGCTCTTCCGCAATATTCTTTTGGAAATGGACTCCTTACAGATGACACCTTGC	2240
AAGAACTAGGACTAGGGAAGCTGCTAAATCGTTACCTTATTATAGCACTTCTCAATGCCACACCTGGGCC	2310
AGATGTGGTTAAAAAGTGCAACCAGGTAGCAGCATGTCTACCAGAAAAATGGTTTGAAAATCTGCCATG	2380
AGGACATCTATTCCACAGCTAGAAAACCTTATTAGTTTATTGTCAGTCTGCACATAAATTATCTAGAA	2450
GTGAATTCAGGGATGAAGTCGAAGAAATAATTCTTATTTTGGTGAATAAAGATTGAATAAATCTTATT	2520
ATCCTTCATAGGAGAGCATCACCTAGACCATCTTAAATCACTAATTAAAGAAGATTGAATAAATCTTATT	2590
GGAAAATGCTAAAAATTTTAAATATAGTTACACTCAGTTCTTTGTTTGAAGAAGAGCTGGTGCTCTCTCT	2660
TCCTTATTCCTGTAAATAGAAGGTAGGATTTGAAAAAAGCAGGACTCCACCTCTGTATTCCCCCGTGCT	2730
TTACCTTCTGGCATCATGAAAAGCTGCCATGATTCTGTGGTGTCTAAGGAATTAATGCACTGGAGCTT	2800
TAAGAGCTCAACGTGTTTCCCTTTGAAAAAAAAAAAAAAAAAAAA	2845

Figure 12

THIS PAGE BLANK (USPTO)

## GCF1 DEDUCED AMINO ACID SEQUENCE

MKKRVTNRRHWTTHRRRQTRKKKKKKRVLGRRALGPRPWLTKRKGFLF 50  
GSARLIPATAMAPRSRLLSLGRRGNFRSRVLRKSRPLEEAARRWRDCPT 100  
GEGALVAGAGSGRAPGVPPKRLPARTKAQNPEPLMCPQMKRKYITPQKV 150  
RMIRVCLLTVLALLEKKNFHQQLRSQMLLFRQPAENVELARAQDDYISL 200  
DVQHTSSISVSRNEETSEESQDEKQDTWEQQQMRKAVKIIERDIDLSC 250  
GSGSSKVKKFDTSISFPFVNLEIISKQLNTRLTLQETHRSHLREYEKYV 300  
QDVKSSKSTIONLESSNQALNCKFYKSMKIYVENLIDCLNEKIINIQEI 350  
ESSMHALLLKQAMTFMKRRQDELKHESTYLLQQLSRKDETSTSGNFSVDEK 400  
TQWILEEIESRRTKRRQARVLSGNCNHQEGTSSDDELPSAEMIDFQKSQG 450  
DILQKQKVFEEVQDDFCNINILLKFQQWREKFPDSYYEAFISLCIPKL 500  
LNPLIRVQLIDWNPLKLESTGLKEMPWFKSVEEFMDSSVEDSKKSSSDK 550  
KVLSAIINKTIIPRLTDFVEFLWDPLSTSQTTSLITHCRVILEEHSTCEN 600  
EVSKSRQDLLKSIVSRMKKAVEDDVFIPLYPKSAVENKTSPHSKFQERQF 650  
WSGLKLFERNILLWNGLLTDDTLQELGLGKLLNRYLIALLNATPGPDVVK 700  
KCNQVAACLPEKWFENSAMRTSIPQLENFIQFLQLQSAHKLSRSEFRDEVE 750  
EIILILVKIKALNQAESFIGEHHLDLKLSLIKED 784

THIS PAGE RELEASED (USPTO)

14/20

# DNA SEQUENCE OF THE EPIDERMAL GROWTH FACTOR RECEPTOR GENE PROMOTER

CCACCGTGTCACCGCCTCGCGGCCGCTGGCCTTGGGTCC -382  
CCGCTGCTGGTTCTCCTCCCTCCTCCTCGCATTCCTCCTCC -342  
TCCTCTGCTCCTCCCGATCCCTCCTCCGCCGCCTGGTCCC -302  
TCCTCCTCCCGCCCTGCCTCCCGCGCCTCGGCCCGCGCGA -262  
GCTAGACGTCCGGGCAGCCCCCGGCGCAGCGCGGCCGCGAG -222  
CAGCCTCCTCCCCCGCACGGTGTGAGCGCCCGCCGCGCC -182  
GAGGCGGCCGGAGTCCCGAGCTAGCCCCGCGGCCGCGCC -142  
GCCCAGACCGGACGACAGGCCACCTCGTCGCGTCCGCCCCG -102  
AGTCCCCGCCTCGCCGCCAACGCCACAACCACCGCGCACG -62  
GCCCCCTGACTCCGTCCAGTATTGATCGGGAGAGCCGGAG -22  
CGAGCTCTTCGGGGAGCAGCG -1

Figure 14

BEST AVAILABLE COPY

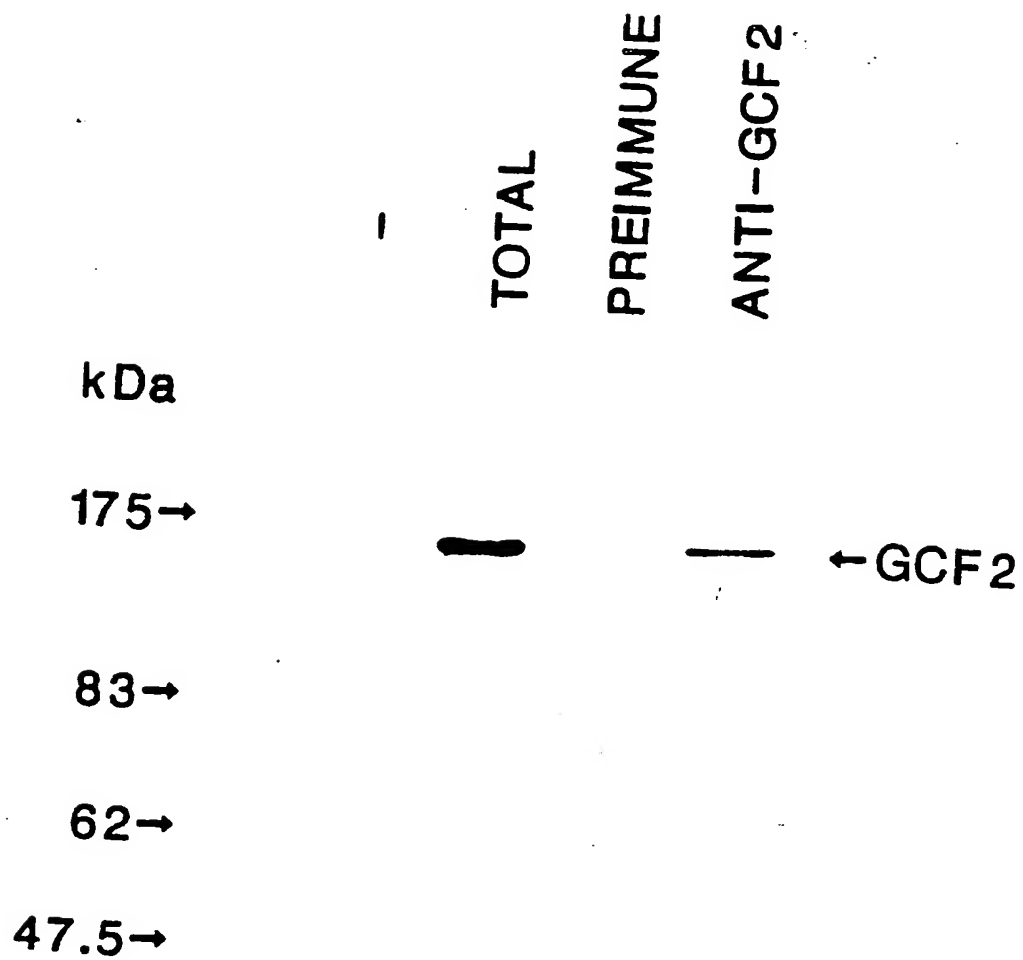


Figure 15



BEST AVAILABLE COPY

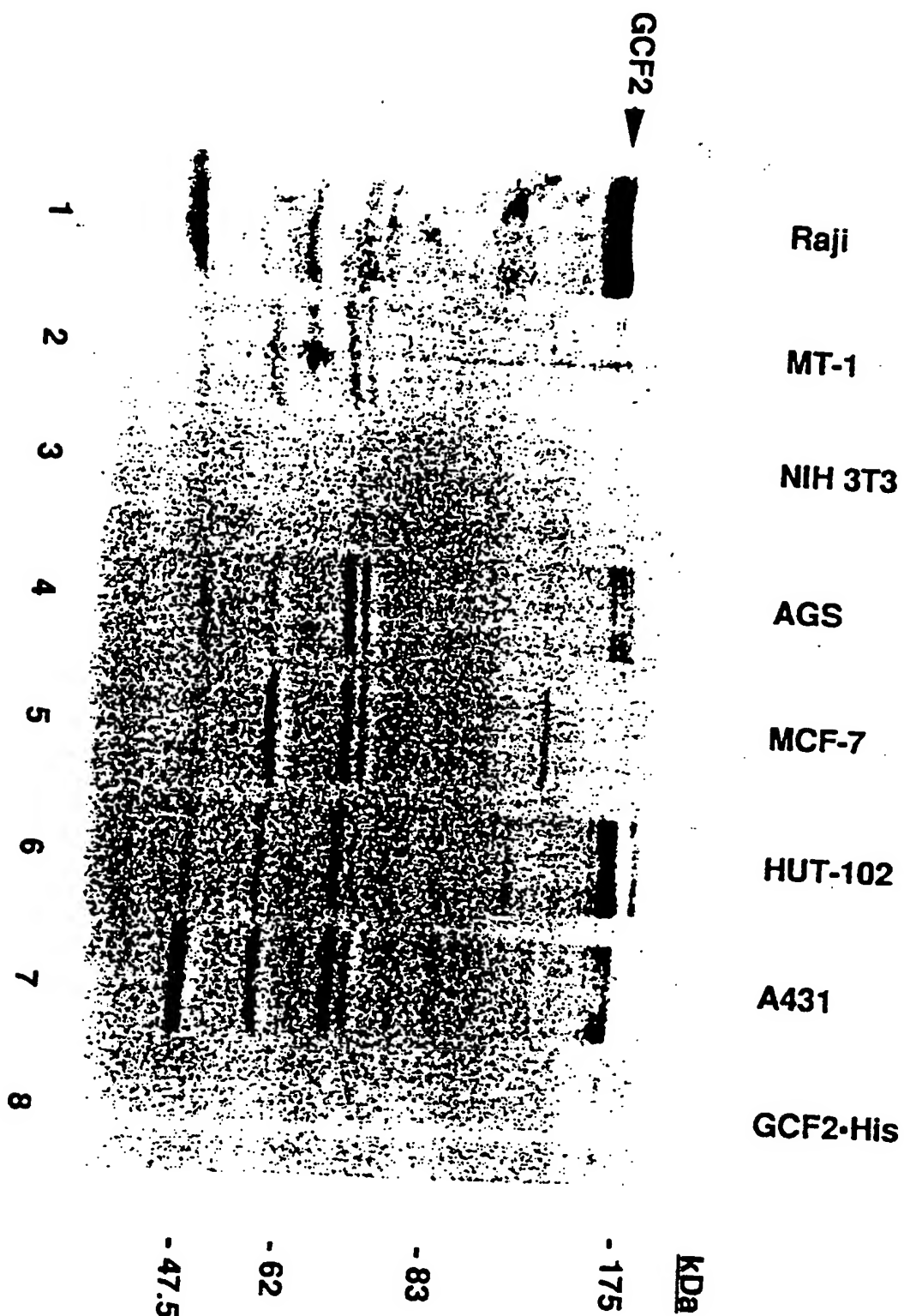


Figure 16

BEST AVAILABLE COPY

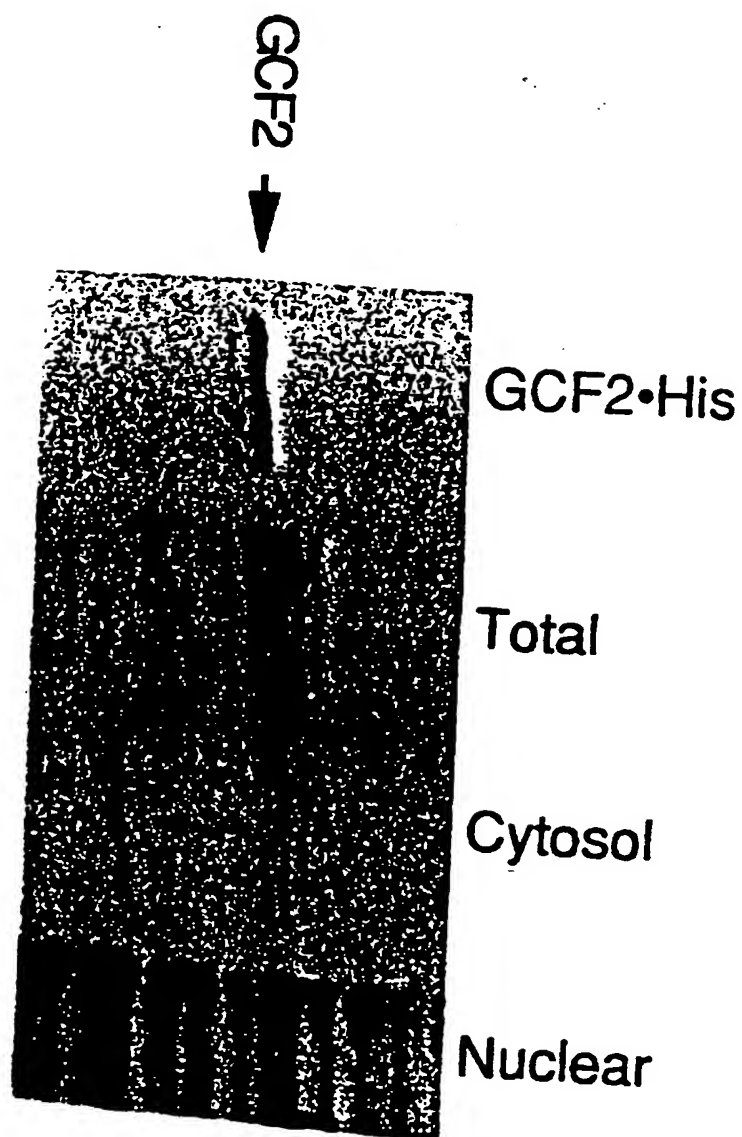


Figure 17

18/20

BEST AVAILABLE COPY

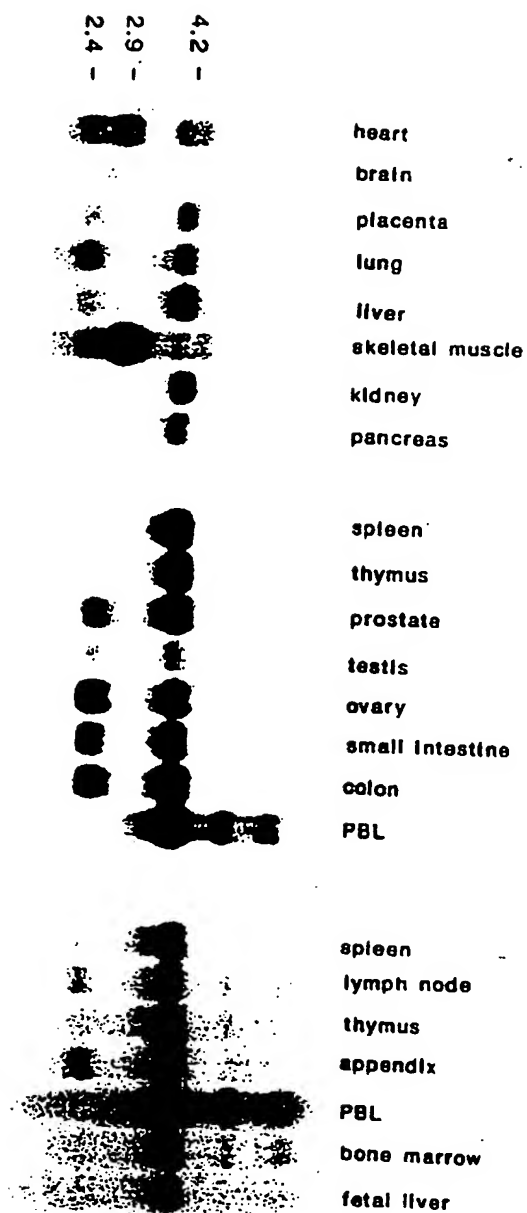


Figure 18

19/20

BEST AVAILABLE COPY

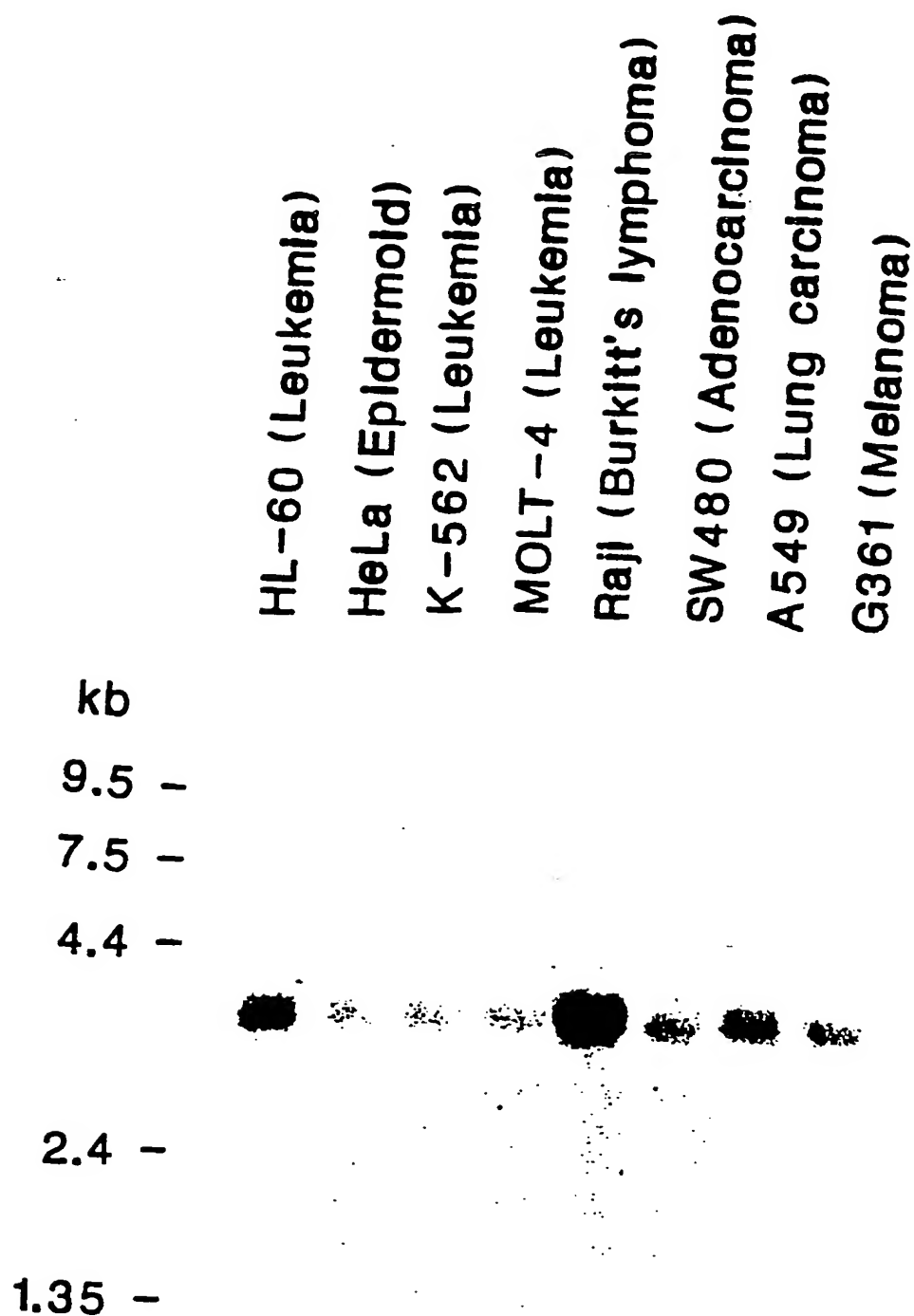


Figure 19

20/20

BEST AVAILABLE COPY

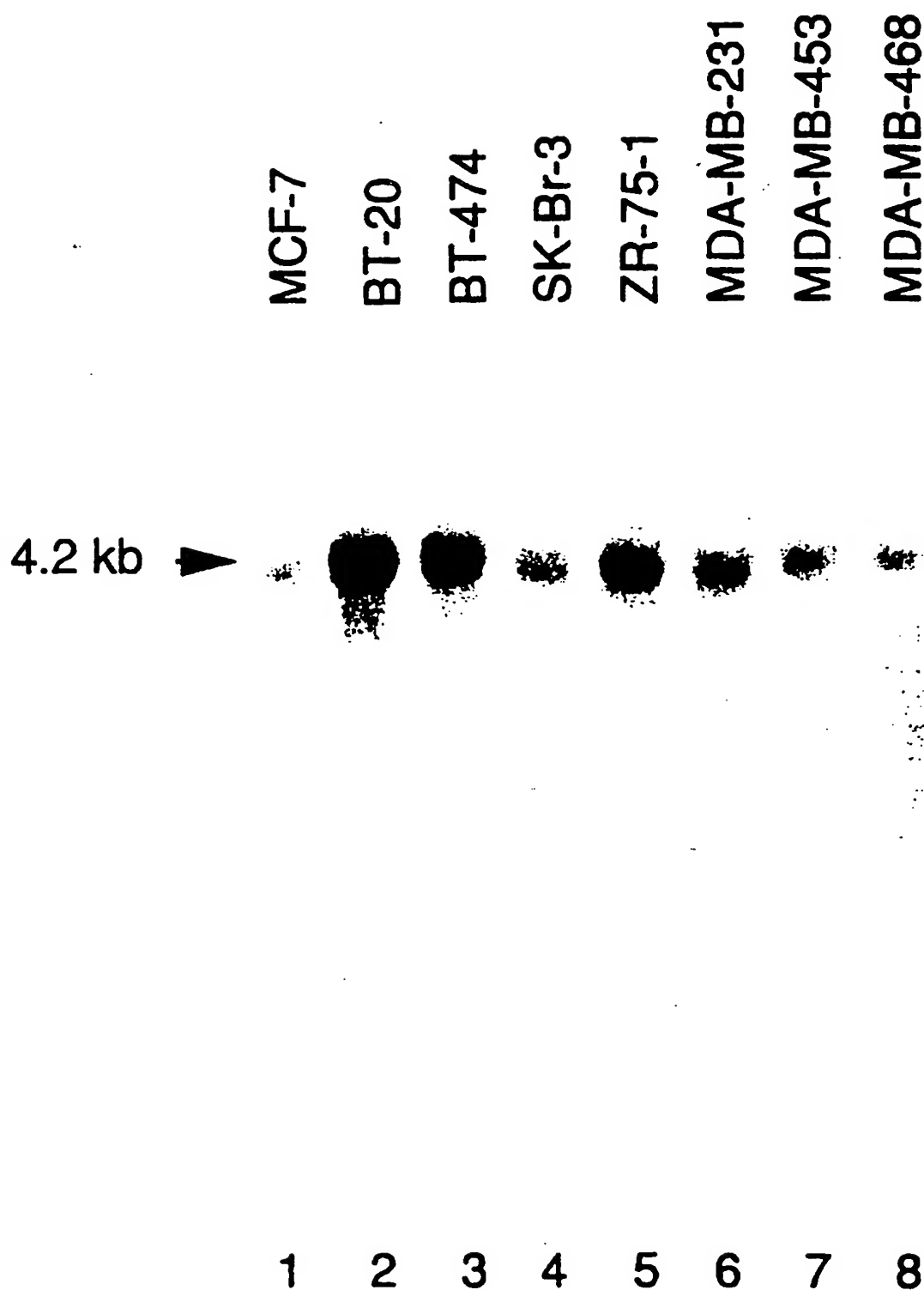


Figure 20

